7/25/68

MODEL AND SOLUTION
OF A LARGE-SCALE, COMPLEX DISTRIBUTION PROBLEM

A THESIS

Presented to

The Faculty of the Division of Graduate

Studies and Research

by

David Michael Miller

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Industrial Engineering

Georgia Institute of Technology

September, 1970

MODEL AND SOLUTION

OF A LARGE-SCALE, COMPLEX DISTRIBUTION PROBLEM

Approved:

Chairman

Date approved by Chairman: 9/10/70

ACKNOWLEDGMENTS

It is my privilege to acknowledge the encouragement and guidance received during the course of this research from my advisor, Dr. David Fyffe. I am also indebted to Drs. Douglas Montgomery and Lynwood Johnson for their critique of the research and their helpful comments.

A word of gratitude must be extended to my wife, Jennie. Without her almost unending patience and her typing skills, this research might never have reached a successful conclusion.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

LIST OF TABLES

LIST OF FIGURES

SUMMARY

The first objective of this research was to formulate a model of a specific large-scale, complex distribution system. This system can be described as a multi-commodity, multi-mode network of plants, warehouses, and customers. A second objective was to develop an effective solution procedure for the model.

The underlying problem situation can be described as the determination of the optimal number, size, and location of distribution warehouses in order to satisfy the demand for several individual products to all customers located in a nation-wide pattern. In addition, the production facility and the transportation mode for resupplying each warehouse, as well as the transportation mode for supplying each customer-product sink must be determined. The solution must be such that the total distribution and warehouse operating costs are minimized while ensuring that a customer service time limit is met in supplying all customer orders.

Several novel concepts are introduced to facilitate the formulation of the model. First, "warehouse location zones" and "customer zones" are utilized as pseudo-warehouses and pseudo-customers. The primary purpose of these zones is to reduce the size and data requirements of the model. Second, a factor to correct for the understatement of warehouse operating cost is added to the model. In the form of a "product multiplier," this correction factor is required when product lines are used in the model to represent individual products.

Incorporated into the model are several assumptions regarding the system being represented. These are

(1) all production-related activities are fixed;

(2) a simple "EOQ" inventory policy is used by all warehouses;

(3) warehouse capacities are unlimited;

(4) customer demand patterns are relatively homogeneous and mutually independent, and are deterministic in nature;

(5) tariff rates along resupply transportation paths are the least-cost rates in the tariff schedule;

(6) the safety stock and reorder point levels at all warehouses are linear functions of annual warehouse throughput;

(7) the calculation of inventory operating costs can be based on the unit production costs of the plant which is the "least-cost" supply plant for a particular warehouse.

The formulation of the model results in a nonconvex, discontinuous, mixed-integer zero-one programming problem. A solution procedure is developed to determine a locally optimal solution. External analyses are utilized to determine the appropriate supply and resupply modes to be used. In a similar manner, the customer service limit restriction is used as an external constraint on data entering the model.

The solution procedure is a two-phase successive linearization algorithm. This procedure is an extension of the basic linearization-solution schemes presented by Baumol and Wolfe and by Hammond. The basic scheme is modified in the two-phase procedure by a) increasing the number of searches of the solution space, b) using an additional "stopping" criteria, c) including a procedure to remove the possibility of suboptimization due to cost curves "crossing," d) using a heuristically derived improved starting point, and e) attempting to identify and select the global optimum utilizing a heuristic penalty function.

An example of the underlying distribution problem is modeled and solved. This problem is similar to the example posed by Baumol and Wolfe. The results presented indicate the usefulness of the second phase of the solution procedure.

One of the major conclusions of the research is that "warehouse location zones," "customer zones," and "product multipliers" are effective modelling techniques for large-scale distribution problems. Also, it was concluded that a local optimum of the nonconvex distribution problem can be obtained through the application of the two-phase successive linearization solution procedure.

CHAPTER I

INTRODUCTION

Many industrial firms are confronted by the problem of optimal design of large scale distribution systems. Such systems have many facets and components, few of which are mutually independent. Among these are the locations of storage facilities, the proper assignment of customer demand, the choice among alternative methods of warehousing, and the use of alternative transportation modes. The optimal design of such a system is often called "the distribution problem." It is this problem that is the subject of the current research.

Description of Problem

The motivation for this research stemmed from the confrontation of a practical distribution problem. The problem involves determination of that system which will meet the annual distribution requirements for some forecasted time period and yet will minimize all distribution costs involved. These distribution requirements are primarily that all customer demand must be met, and that an imposed service or delivery time limit must not be exceeded in supplying customer orders. The components of the required system are the number, location and operating mode of all warehouses in the system, the supply and resupply transportation modes to employ, and the proper assignment of customer demand. A concise statement of this problem is given in the following

GIVEN: (1) a set of customers with forecasted monthly product demand data (tonnage);

(2) a set of production facilities (plants), some of which may have fixed production levels (capacity);

(3) three available modes of transportation (truck, rail, piggyback) and tariff rates schedules for any point-to-point shipment;

(4) warehouse costs for each of the alternative warehouse types, Public or Lease;

FIND: That distribution system which minimizes total transportation and warehouse operating cost, and consist of the following parameters:

(a) the number of warehouses to use;

(b) the location of each warehouse;

(c) whether each warehouse should be leased or rented;

(d) which warehouse(s) to use to supply the tonnage demanded by each customer region for each of three product lines;

(e) the total tonnage to be shipped through each warehouse (and, hence, its size);

(f) the plant to use to resupply each of the warehouses;

(g) the transportation mode to use to resupply each warehouse from its resupply plant, and to ship customer demand (tonnage) from the warehouse(s) which supplies it (under normal operating conditions);

CONSTRAINTS: These parameters must be determined so that the following conditions are met:

(i) each customer's demand for each of the three product lines in 1974 must be met;

(ii) all customer orders must be delivered within a three day time span after warehouse receipt of the order;

(iii) the production level at pertinent plants must not be increased.

Purpose of Research

The objective of this study is to formulate an appropriate model for the complex large-scale distribution problem just described and to develop an effective solution procedure. The model should be a realistic representation of the physical distribution system under study. However, in order to have merit, this model must be adaptable to effective solution techniques.

The solution procedure developed should produce "good" results, and should be capable of being justified under appropriate criteria.

The problem associated with solving the model developed in this paper constitutes the minimization of a strictly concave function over a convex feasible region. The theoretical difficulties associated with this problem, along with the complexities involved in modeling a large-scale distribution system, constitute a valid and interesting research assignment.

Scope of Research

It should be recognized that there are several different solution approaches that can be taken to solve the model developed. The current research is limited to applying only one of these approaches. Hence, no comparison of alternative approaches will be made to establish the relative efficiency or effectiveness of the solution method used.

Also, the research will be limited to the formulation and development of a solution procedure, with no effort devoted to improving the techniques and ideas employed. For example, a penalty function is utilized in the solution procedure and is based on a subjective criteria for improving a suboptimal solution. If time were available, the "best" formulation of this penalty function should be determined.

Method of Procedure

The first section of the paper is devoted to reporting a survey of published literature which is relevant to location problem modeling and solution procedures. This section establishes the background necessary for an understanding of the problem area being investigated. The basic formulation of location models are presented and discussed. Solution procedures applicable to the various model formulations are presented, and the advantages and disadvantages of each approach are indicated.

The next section describes the physical distribution system under study. A model of this system is first presented in general terms and used as a guideline to explain the various components and parameters which are incorporated into the model. After these components are described and the basic assumptions underlying the formulation and application of the model are presented, a specific, quantitative version of the model is developed. The form of this model is a strictly concave, discontinuous minimization problem.

The theoretical difficulties of this type problem are presented in the next section, along with the solution procedure which was developed to "solve" this problem. Although an exact solution is impractical, a two-phase method is described which will lead to a "good" solution. This two-phase method is a successive linearization-solution procedure. However, it is a modification, based on the recognition of several theoretical difficulties inherent in ordinary linear approximate solutions to nonlinear problems.

A justification of the two-phase procedure is given which is based on the fact that a) the procedure "efficiently" produces "good" results, and b) the procedure improves upon similar successive linearization approaches presented in the literature.

The fifth section of the paper presents the results of applying the model and solution procedure to an example distribution problem similar to those encountered in industry. The specific problem, data base, and solution results are given.

The final section is devoted to summarizing the results previously established and to drawing conclusions from the research effort. Several possible refinements of the model and solution procedure are indicated. Also, areas are given into which further research effort may be profitably extended.

Limitations of Approach

There are several assumptions and characteristics of the model and its solution procedure which may limit their effectiveness in certain applications. These are:

Model

1) the model parameters are deterministic in nature,

2) in order to accurately represent individual customers in conglomerate zones, customer demand characteristics should be fairly uniform,

3) formal inventory strategies exist at all storage facilities.

Solution Procedure

1) requires several post-solution sensitivity analyses to obtain best results,

2) results are not <u>necessarily</u> globally optimal,

3) the procedure does not emphasize maximum efficiency of computational speed.

CHAPTER II

BACKGROUND

Introduction

Among the most important problems facing corporate-level decision-makers is that of distributing planning. Not only does this area rank as one of the largest contributors to the final cost of goods and services sold, but the potential cost savings in this area are normally very large. For example, Atkins [1] reports that Standard Oil of New Jersey recently attributed annual savings of $5,000,000 to the use of a single mathematical model for planning distribution and manufacturing operations. Similarly, another major corporation reported that they reduced distribution cost from nine per cent to six per cent of total cost of goods sold.

Distribution planning is also one of the most complex and difficult problems confronting decision-makers. There are many facets of logistics which must be considered in planning and analyzing distribution systems. Transportation mode alternatives, product demand patterns, intermediate storage desirability, and product delivery requirements (such as a lead time or service time requirement), are all influences on the distribution system. Planning must be based on both subjective and objective analysis of these influences or system components. Unfortunately, the cost structure underlying objective analysis of a distribution system is generally difficult to handle once it has been modeled.

The basis of most distribution analyses is what is often labeled the "warehouse location" problem. Essentially, this is the problem of determining the least-cost network (number, location, and type) of intermediate storage facilities to use for supplying customer demand from a set of plants or sources.

The modelling of such networks can become very complex when many relevant influences are explicitly considered. The availability of alternative modes of transportation and of alternative types of warehousing facilities are examples of such consideration. The general warehouse location model can be expressed as, letting

X_{ijm} = quantity of product m shipped from plant i to warehouse j,

X_{ikm} = quantity of m shipped from i to customer k,

X_{jkm} = quantity of m shipped from j to k,

T_{ij} = total quantity shipped from plant i to warehouse j,

T_{ik} = total quantity shipped from i to k,

I_{jk} = total quantity shipped from j to k

W_j = total throughput of goods through warehouse j,

CAP_{im} = capacity of plant i to produce product m,

CAP_j = capacity of warehouse j to store goods shipped through it,

$f(T)$ = a function of T,

$$\text{minimize } Z = \sum_i \sum_j f_{ij}(T_{ij}) + \sum_i \sum_k f_{ik}(T_{ik}) \tag{2.1}$$

$$+ \sum_j \sum_k f_{jk}(T_{jk}) + \sum_j f_j(W_j),$$

subject to:

$$T_{ij} = \sum_m X_{ijm} ,$$

$$T_{ik} = \sum_m X_{ikm} ,$$

$$T_{jk} = \sum_m X_{jkm} ,$$

$$W_j = \sum_k \sum_m X_{jkm} ,$$

and,

$$\sum_i X_{ikm} + \sum_j X_{jkm} = d_{k,m} , \text{for all } k, m, \quad \left\{ \begin{array}{l} \text{All customer} \\ \text{demand must be} \\ \text{satisfied} \end{array} \right.$$

$$\sum_i X_{ijm} = \sum_k X_{jkm} , \text{for all } j, m, \quad \left\{ \begin{array}{l} \text{Total flow of each pro-} \\ \text{duct into each warehouse} \\ \text{must equal total flow out} \end{array} \right.$$

$$\sum_j X_{ijm} + \sum_k X_{ikm} \le CAP_{im} \quad \text{for all } i, m, \quad \left\{ \begin{array}{l} \text{Capacity of each plant} \\ \text{for producing each pro-} \\ \text{duct must not be} \\ \text{exceeded} \end{array} \right.$$

$$\sum_k \sum_m X_{jkm} \le CAP_j \text{ for all } j. \quad \left\{ \begin{array}{l} \text{Total throughput of} \\ \text{each warehouse must} \\ \text{not exceed capacity} \end{array} \right.$$

Problem Formulation

Although most models of real-world distribution systems represent transportation cost as a linear function and warehousing cost as a concave function -- $f_j(W_j)$ in the previous model--the underlying cost structure of both components are normally concave in form. Transportation cost in a physical distribution system are based on the transportation carrier's tariff rate structure for any point-to-point shipment. These tariff rates are a function of the amounts of goods shipped, the unit rate decreasing at specific volume or lot break points. Therefore, the transportation cost of point-to-point shipments is a piece-wise, concave-shaped function of the amount shipped (see Figure 4). Also, specific rates are based on such factors as the crossing of state boundaries, the terminal point of the shipment being a terminal "head" for the carrier in question, and other such considerations. For this reason, a shipment of the same amount of goods the same distance from the same starting point may not result in the same transportation cost due to the location of the terminal shipment point. Normally, the characteristics of transportation rate structures and costs are the same regardless of the carrier or transportation mode.

The actual cost of operating warehouses in the distribution system is a function of the amount handled by the warehouses. Specifically, the unit cost of handling and storing goods is a decreasing function of the volume throughput of a warehouse. When the warehouse in question is a public warehouse, these economies of scale associated with operations at larger volumes are due to the possibility of renegotiating storage and handling charges (rates). Also, economies of scale are due to the use of quantitatively derived inventory reordering strategies which result in a less than proportional increase in inventory level with an increase in demand or throughput. The mathematical model of the general form of this inventory order quantity relation is,

$$Q = aD^n \quad \text{for} \quad 0 \leq n \leq 1 \tag{2.2}$$

where Q = reorder quantity
D = annual demand assigned to the warehouse
n = a constant
a = a constant

Since warehousing cost are a function of inventory levels, they are concave in form. An example of an inventory strategy causing these concave costs is the simple "EOQ" policy in which $n = \frac{1}{2}$.

The cost of operating a leased type warehouse are also a concave function of the throughput level of the warehouse. As in the public warehouse case, rates can be renegotiated with larger storage area requirements, and economics of scale also result from use of certain inventory policies. However, increases in throughput volume force use of additional handling equipment and personnel at certain throughput levels. Although this tends to cancel out the reduction in leasing charges at break-point levels, the warehouse cost function is still concave due to the results of inventory policies. Company-owned warehouses have cost functions which are similar in nature to those of leased warehouses.

The concavity of the two components of the total distribution cost function (transportation and warehouse operating cost) of model (2.1) cause an interesting yet difficult computational problem. Essentially, this is the minimization of a concave function over a convex set of constraints. The values of the decision variables (within the bounds of the problem constraints) are being sought which cause the concave-shape objective function to have a least-cost solution, or its smallest value. The difficulty in solving this problem is that there exist many "local optima," solutions which are least-cost solutions within small ranges of the decision variables--and one "global optimum" or best solution. Most mathematical solution techniques are designed to search the space of possible solutions to the problem and find a solution which cannot be improved upon within a local area. The underlying assumption in these techniques is that any local optimal solution is also the global optimum. That is, that there exist only one solution which cannot be improved upon.

In a two-dimensional problem, it can easily be seen that the least-cost solution lies at one of the end-points of the constrained function and that both end-points may in fact be local optima (see figure 1). A graphical interpretation of the three-dimensional problem is given in figure 2. In an n-dimensional warehouse location problem (with m constraints), Balinski [2, p. 287] has observed that there are m^n extreme-point solutions when the capacity limitations are relaxed. Even with capacity constraints the number of extreme points approaches m^n. In problems having a cost function of a general concave form, the only solution technique capable of producing the global optimum is complete enumeration of all extreme points. Obviously, this approach is impractical for large problems.

However, more efficient solution techniques can be utilized when varying degrees of generality are used in representing the problem's cost equation. Each of these formulations still results in a concave function. The simplest formulation is that of the so-called fixed-charge problem. Here there is a constant or fixed cost associated with operating the facility in question, regardless of the volume handled by the facility. In addition, there is a variable cost which is a linear function of the volume handled (see figure 3).

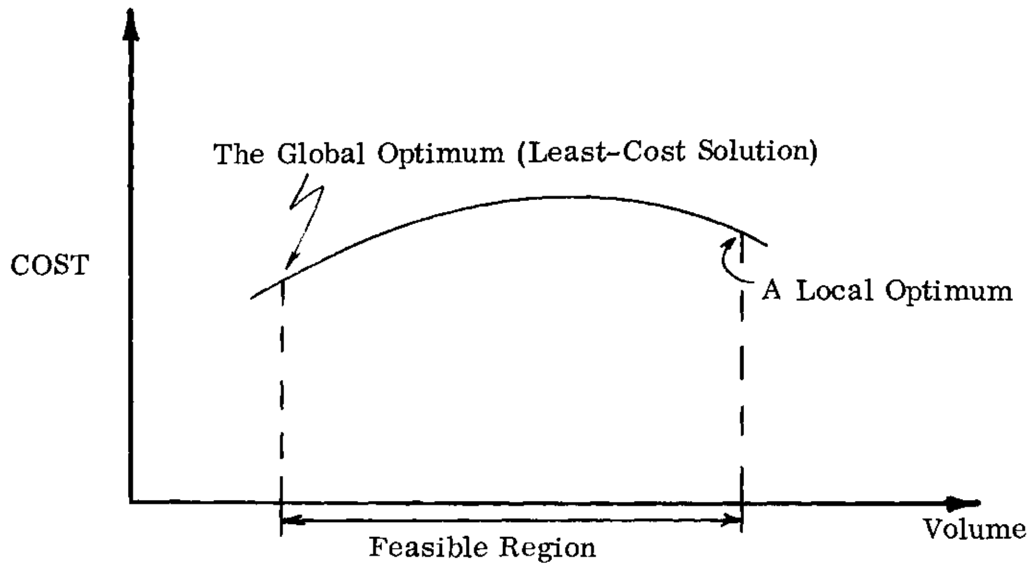


Figure 1. Minimizing a Concave Function (Two Dimensions).

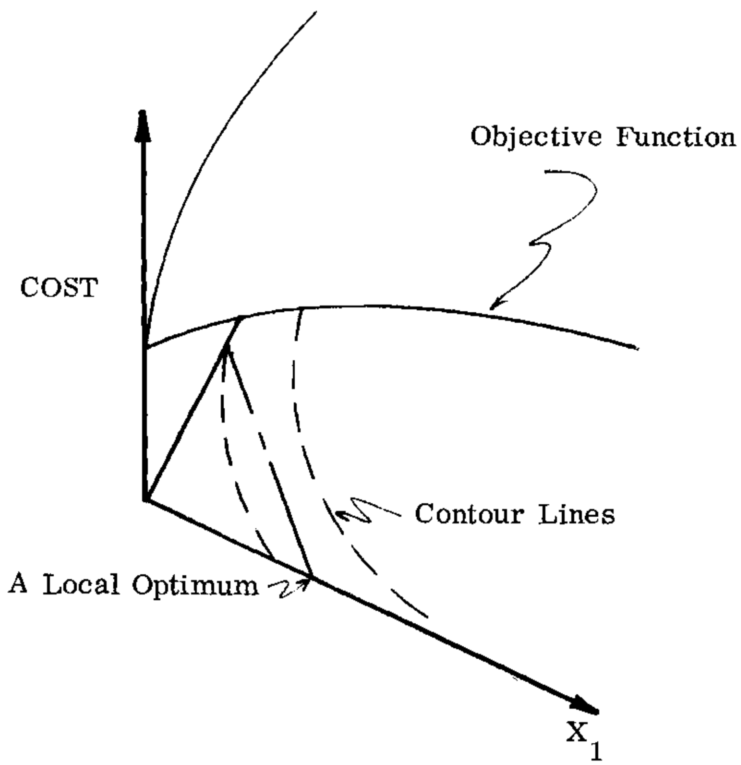


Figure 2. Minimizing a Concave Function (Three Dimensions).

A somewhat more realistic representation is that of the piece-wise linear formulation (see figure 4). In essence, this is a series of fixed charge problems (see figure 5). There may or may not be an initial fixed cost component, but the unit charge is not constant over the entire volume range. That is, there are economics of scale associated with larger volumes. These economics are represented by smaller unit cost in larger volume intervals.

The most general formulation of the cost function is that of a differentiable, strictly concave function with discontinuities (see figures 6, 7). In this case, there may or may not be an initial fixed cost. There are break-points or discontinuities in the function due to discrete jumps in one or more cost components. The continuous portions of the function are characterized by increasing total cost but at a continuously decreasing rate.

The specific cost formulation used in a warehouse location problem dictates the most applicable solution approach. There are four broad categories into which most solution approaches fall. These are mathematical programming approaches, simulation, successive linearization, and heuristics approaches. When the cost function has been modeled as a piece-wise (or fixed charge) function, mathematical programming methods are most applicable due to the use of zero-one decision variables in formulating fixed-cost components and the linearity of the continuous intervals of the function. However, when these continuous intervals are nonlinear, one of the other approaches is generally a more useful and effective solution tool.

The remainder of this chapter is devoted to investigating the literature

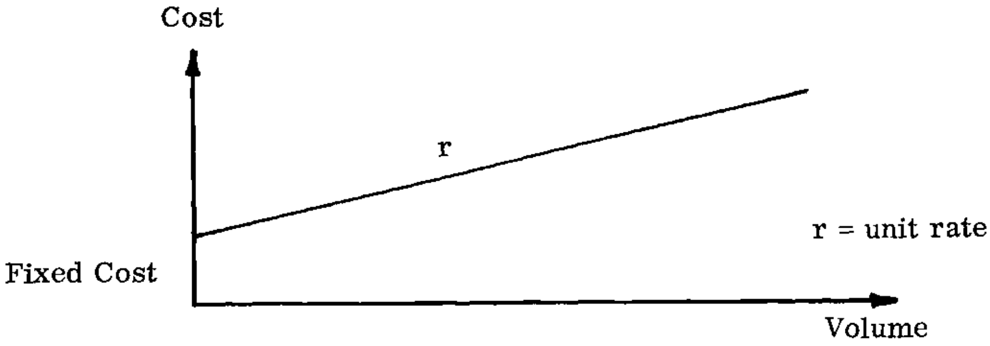


Figure 3. Fixed-Charge Function.

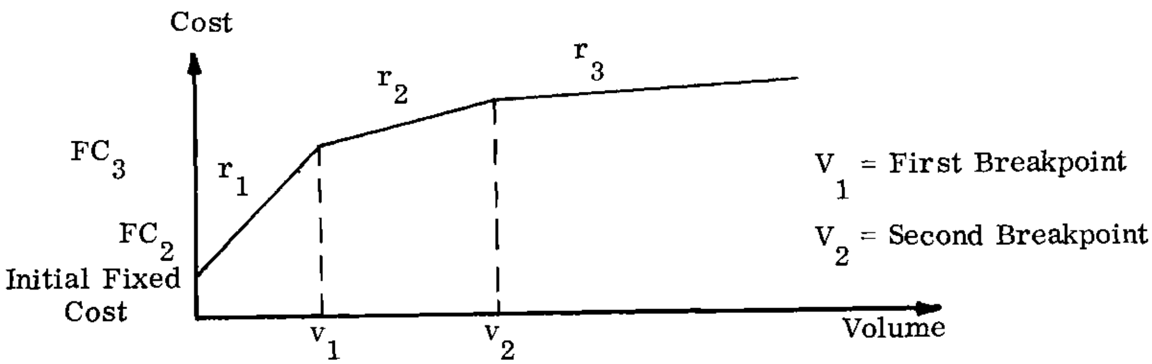


Figure 4. Piece-Wise Linear Function.

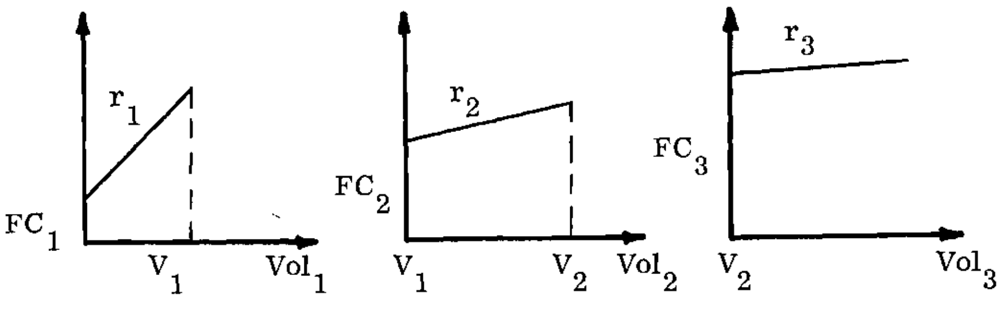


Figure 5. Piece-Wise Linear Function As a Series

of Fixed-Charge Functions.

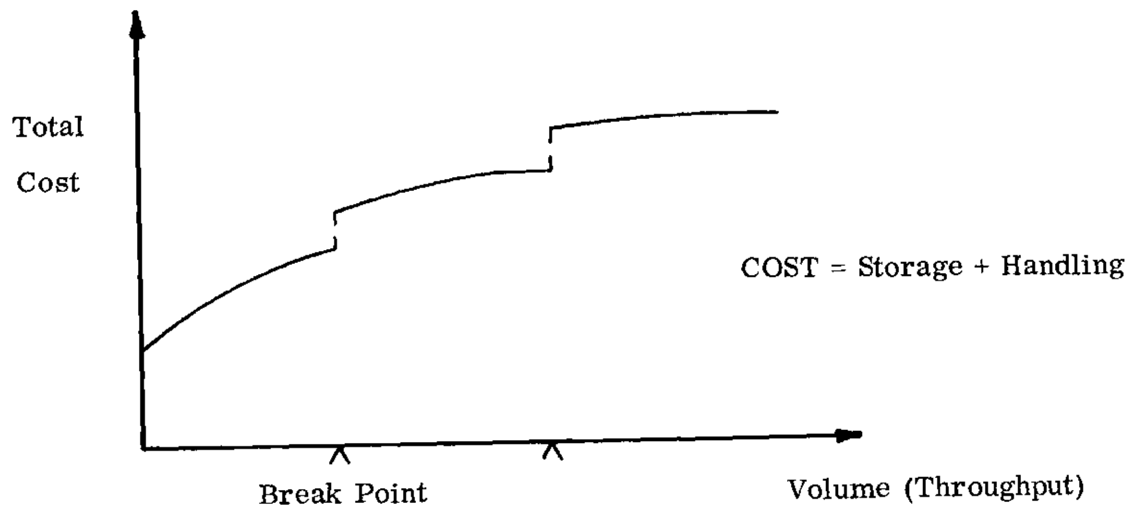


Figure 6. Operating Cost Function for Leased Warehousing.

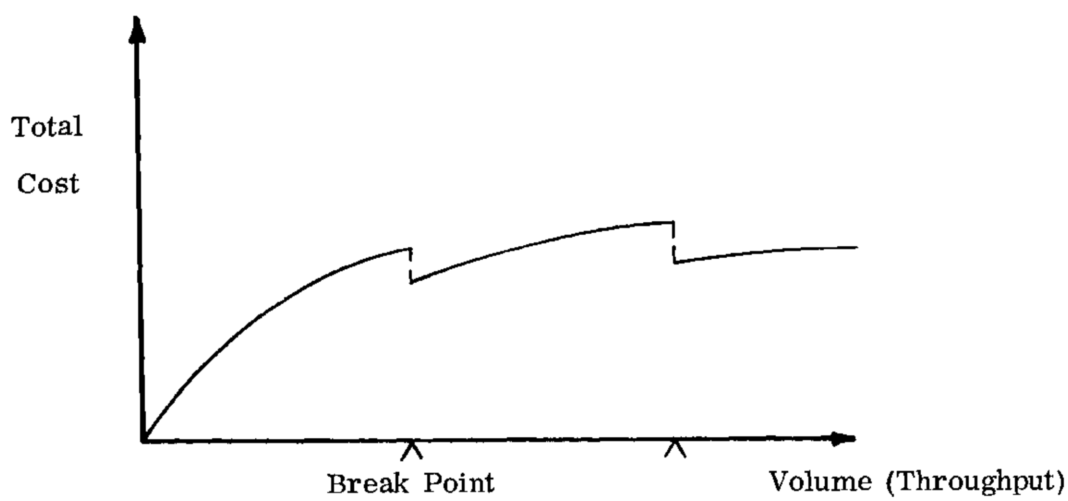


Figure 7. Operating Cost Function for Public Warehousing.

that has been reported concerning each of these solution approach categories. Such an investigation is necessary in order that the most advantageous model and solution procedure be developed to solve the distribution problem described in the first chapter.

Alternative Solution Approaches

The basic warehouse location problem under study is a special type of the so-called "location-allocation" problem. Most modeling and solution efforts reported in the literature have been devoted to this location-allocation class of problems rather than specific warehouse location problems. The solution approaches applicable to this latter group are a subset of approaches to the more general location-allocation category. Hence, the relevant area of investigation of solution techniques is the literature pertaining to location-allocation problems.

A significant difference between these two classes of problems which should be noted is that in the warehouse location problem costs associated with a particular transportation path (such as shipments from warehouse A to customer B) are dependent on shipments along other paths in the system. However, in some location-allocation problems (such as the "simple plant location" problem) costs along each path are independent of other paths. That is, one or more cost components of the warehouse problem are functions of the type $f_i(\sum_j X_{ij})$, whereas cost functions in other location problems may be of the form $f_{ij}(X_{ij})$. Transportation costs for a path i-j are an example of the former function. Recognition of this difference is of importance in the investigation of modeling and solution procedures. As stated earlier, there are four categories of approaches:

mathematical programming, simulation, heuristics, and successive linearization.

Mathematical Programming

The piece-wise linear formulation requires that a fixed cost be incurred if and only if a facility (or specific interval of a function) is used. The representation of this requirement in terms of zero-one decision variables, and the linear portions of this cost formulation make this particular model very adaptive to mixed integer, zero-one programming. A model of this type formulated by Balinski in 1964 [3] has served as the basis for most modeling efforts in the mixed integer mode. His model, for an uncapacitated location problem, is as follows:

$$\text{minimize } Z = \sum_i \sum_k C_{ik} X_{ik} + \sum_i f_i Y_i , \tag{2.3}$$

$$\text{subject to:} \quad \sum_i X_{ik} = 1, \quad \text{for all } k,$$

$$0 \leq X_{ik} \leq 1 \text{ for all } i, k,$$

$$Y_i = 0, 1 \quad \text{for all } i,$$

where

C_{ik} = total transportation cost from plant i to customer k,

X_{ik} = fraction of customer k's demand supplied from plant i,

Y_i = fixed cost associated with plant i.

Efroymson and Ray [4] noted that Balinski's formulation resulted in a large, inefficient linear program when the Y's are not constrained to be integers (this embedded linear program is inherent in most integer programming solution techniques). Therefore, they reformulated the model so as to have a more efficient embedded linear program and, hence, a model more easily solved for large problems. Whereas Balinski used the partitioning theorem of Benders to solve his model, Efroymson and Ray utilized a branch and bound solution scheme. Incorporated in this scheme are several decision rules used to overcome an inherent disadvantage in their formulation. This disadvantage is that the solution to the embedded linear program will result in a relatively large number of plants being specified as "open" or utilized. Each customer will therefore absorb a high percentage of the fixed cost of the plant to which that customer is assigned. This bias seems to be inherent in any linear programming technique used in the solution of location-allocation problems.

The basic model of Efroymson and Ray is concerned with the simple fixed charge problem (uncapacitated). However, they show how an extension can be made to include the piece-wise linear cost function. The essential requirement that a zero-one variable be associated with each segment or piece of the cost function obviously results in a much larger problem. Solution times associated with large mixed-integer models can often be prohibitively excessive. Storage requirements for node results as well as the computational time involved in embedded linear programming solutions are extremely large for problems of "practical" size. However, Efroymson and Ray report computational results of solving a 200 customer problem involving 50 zero-one variables in approximately 10 minutes (their computer code utilized several "short cuts," such as starting with an a priori "good" solution).

A more recent mixed-integer formulation by Sa′ [5] includes capacity limitation of the plants to be located, a restriction not included in the Balinski or Efroymson and Ray approaches. Sa′'s mathematical model is similar to the Efroymson and Ray reformulation of Balinski's basic model (2.3) with the addition of the plant capacity constraints. Branch and bound is the solution technique used by Sa′. This branch and bound procedure produces, as do other mixed-integer solution approaches, the global optimum solution. However, as pointed out previously, large problems cannot be solved efficiently using this solution mode. Admitting this shortcoming, Sa′ solves the same model with a heuristic or approximate algorithm. Consisting of two phases, this approximate routine is a combination of "dropping" locations from consideration and "adding" location to some starting subset of locations. Phase one results in a "good" but possible suboptimal solution. Phase two, as Sa′ proves, will result in the global optimum. His second phase is essentially a reexamination of extreme points not considered by phase one. That is, if the decision maker is willing to accept the fact that the solution may not be the best possible solution, he needs only to utilize phase one of the algorithm. If, on the other hand, he desires to find the best answer, phase two (along with extra solution effort) can be undertaken.

Heuristics

Sa′'s approximate algorithm falls in the class of location-allocation solution approaches characterized as heuristic techniques. These approaches utilize decision rules based on economically sound principles or heuristics. The validity of these principles may not be supported by formal proofs. Therefore, these heuristic approaches seek the optimum or best answer to the problem but do not guarantee that the final answer will in fact be a global result. An important point to realize is that a heuristic algorithm may produce the best or global optimum solution rather than just a good or local optimum answer, but no guarantee is given of this.

As previously indicated, heuristic methods usually are derived to solve the general location-allocation cost formulation rather than the more restrictive piece-wise linear formulation. This fact offers an obvious advantage in utilizing heuristic as opposed to, say, mixed integer programming in formulating and solving actual location-allocation problems. That is, a more realistic, true-life picture can be represented in the model to be solved by heuristics. Also, heuristic methods are not usually as limited as math programming methods in the size of problems that can be solved. An upper limit on the number of zero-one variables (fixed cost components) in the cost formulation of problems which can efficiently be solved by mixed integer programming is usually acknowledged to be 100 such variables [6, p. 844]. Of course these advantages are countered by the fact that heuristic approaches generally do not guarantee globally optimum results.

In 1963 Kuehn and Hamburger [7] developed a heuristic program to solve a warehouse location-allocation problem. The multi-product model which they formulated was based on the fixed charge interpretation of warehouse operating cost. However, Feldman, Lehrer, and Ray [8] extended Kuehn and Hamburger's model to include the more general cost formulation of a differentiable, strictly concave function. Their model, considering just a single product, and ignoring capacity limitations is

$$\text{minimize} \quad Z = \sum_i \sum_j \sum_k C_{ijk} X_{ijk} + \sum_j f_j(\sum_i \sum_k X_{ijk}) \,, \tag{2.4}$$

$$\text{subject to:} \qquad \sum_i \sum_j X_{ijk} = D_k, \text{ for all } k,$$

$$X_{ijk} \geq 0, \text{ for all } i, j, k,$$

where

X_{ijk} = flow from plant i to warehouse j to customer k,

C_{ijk} = unit cost of flow X_{ijk},

D_k = demand of customer k,

$f_j(\sum_i \sum_k X_{ijk})$ = warehouse j's operating cost as a function of total flow through j, a concave function.

In the Kuehn and Hamburger model, the warehousing cost is represented as

$$f_j(\sum_i \sum_k X_{ijk}) = \begin{cases} a_j + b_j \, [\sum_i \sum_k X_{ijk}], & \text{for } \sum_i \sum_k X_{ijk} \neq 0 \\ 0 & \text{, otherwise} \end{cases}$$

where

a_j and b_j are constants.

It is interesting to note that in the paper published by Feldman, Lehrer, and Ray the "i" subscript is omitted in the mathematical model presented. That is, there is no apparent consideration of the transportation path between plants (i) and warehouses (the i in their model refers to warehouses). However, in reporting computational results, the authors solve a problem involving shipments from plants to warehouses, indicating that the subscript omission may be an oversight.

The heuristic program of Feldman (and Lehrer and Ray) proceeds by first selecting a warehouse flow throughput or reference level for each warehouse. The initial reference level for a warehouse is constructed from the set of customers who are closest to that warehouse on a transportation cost basis. This set is defined as a warehouse's "local customer set." The warehouse is said to have a "local volume" that is the sum of the individual demands of its local customer set. The relative magnitude of each warehouse's local customer set is a preliminary measure of the extent to which the warehouse is centrally located. This magnitude also indicates the relevant decision-making portion of each warehouse's operating cost curve.

The next step in the procedure is to examine the incremental cost of supplying a given customer from each of the available warehouses, assuming that these warehouses have throughput levels equal to their "local volume." Then warehouses are dropped one at a time from a list of potential drop candidates that appear to give cost savings if they were not in the solution. Each time a warehouse is dropped from a previous solution, a new solution or assignment of customers to warehouses is determined utilizing the throughput volumes of the previous solution to calculate marginal or incremental costs. This cycle of dropping warehouses from a list of potential candidates is repeated until the list is exhausted and all warehouses have been examined and eliminated.

An alternative heuristic to dropping warehouses from a starting list is that of adding warehouses which indicate they would produce cost savings if they were in the solution. Kuehn and Hamburger use the latter heuristic as the basis of their location algorithm.

Simulation

Location-allocation problems having the general concave cost formulation have also been attacked through the use of simulation. This approach, as does heuristic programming, produces results which generally are not guaranteed to be optimal. The value of a simulation approach to solving location problems obviously is that much more detail and realism can be incorporated into the model of the problem. Solutions may not be near-optimal but they are based on a more realistic view of the problem. However, simulation models of large, complex systems might easily consume too much time and computer hardware resources, making such a model impractical.

The only simulation model reported in the literature is that of Shycon and Maffei [9]. Although the mathematical model developed is not given in their article, their approach is evidently based on all location sites being specified in longitude and latitude coordinates. This necessitates shipping cost to be approximated by a unit cost multiplied by air miles between any two points. Since tariff rate structures are not singularly correlated with straight-line distances--as indicated previously, factors such as state boundaries, or destination point are often more significant--use of this type transportation cost data may lead to gross inaccuracies. Simulation does, however, allow explicit consideration to be given to variations in customer ordering patterns, order quantities and frequencies.

Such variations are often realistic and significant. Unfortunately, most other programming approaches must rely on averaged or expected ordering factors and consequently may also introduce inaccuracies into the model.

An important component of a simulation procedure to solving a location-allocation problem which is missing in the Shycon and Maffei approach is a means of going from one facility-network to another. That is, a strategy for moving through the list of possible problem solutions is needed. Without an efficient strategy of this type, simulation offers little as a solution technique for large distribution or location problems.

Successive Linearization

Another approach that has been used to solve the location-allocation having a general cost structure is successive linearization. This technique can be defined to be the repeated use of pertinent data to determine the marginal or unit rate associated with a nonlinear function at some value or interval of interest of the function. Applied to the location problem, successive linearization proceeds by 1) approximating the throughput levels of all facilities in question to obtain a linear model, 2) solving the model; and 3) using this solution to revise all throughput levels and successively repeating the procedure. Like the heuristic methods, this approach has the inherent difficulty of potential suboptimal results.

Baumol and Wolfe [10] presented a single-product model incorporating a continuous, differentiable, strictly concave cost function. The mathematical model they investigated is identical to that posed by Feldman (2.4), except for an added warehouse capacity constraint.

In order to transform the model into a standard transportation problem, Baumol and Wolfe replaced the three subscript notation by a two subscript model through the use of a decision rule which specifies that shipments from plants to customers will be made via the warehouse offering the lowest delivery cost. They ignored warehouse capacity constraints (assuming public warehousing space will be unlimited) but apparently added plant capacity limitations.

The iterative procedure utilized by Baumol and Wolfe to solve the original model is given in the following outline.

Initial Stage

1) For each plant to customer path (i to k) determine the warehouse to use as an intermediate stock point based only on transportation cost. Designate the corresponding unit cost as $C^o_{i,k}$. That is,

$$C^o_{ik} = \underset{j}{\text{minimum}} \, (C_{ijk})$$

where the superscript indicates the stage.

2) Solve a standard transportation problem using plants as sources, customers as sinks, and C^o_{ik} as unit cost.

Nth Stage

3) Determine each warehouse's throughput level (sum of customer demand assigned to it) - W^n_j. That is, calculate

$$W^n_j = \sum_i \sum_{k \in J_j} X^{n-1}_{ik} ,$$

where J_j = set of i to k paths using warehouse j as its intermediate storage point.

4) Calculate new unit cost for each i to k path - C^n_{ik} - as follows:

$$C_{ik}^{n} = \underset{j}{\text{minimum}} \left[C_{ijk} + \frac{df_j(W_j^{n-1})}{dW_j^{n-1}} \right] .$$

Final Stage

5) Repeat (2) thru (4) until the throughput levels for all warehouses remain unchanged for two successive iterations. That is, until

$$\sum_i \sum_k X_{ijk}^{n} = \sum_i \sum_k X_{ijk}^{n-1}, \text{ for all } j.$$

Baumol and Wolfe recognized the fact that their procedure may not produce globally optimal results but they proved that it would reduce the total cost (Z) in each step or iteration. They also recognized a bias in solutions resulting from their procedure which is apparently common to all successive linearization approaches. As in the disadvantage reported by Efroysman and Ray discussed previously, solutions tend to incorporate more warehouses than might be a priori expected. The reason for this is that the full extent or impact of the economics of scale or decreasing marginal cost are not considered in the linearization of the concave cost function. In other words, the "bending down" of the cost curve is not fully recognized.

Balinski and Mills [11] also utilized a linearization technique in solving a warehouse location model. However, their approach does not successively update the linearized cost function. Rather, one linearized solution is used as a lower bound on the problem's optimal solution and a second solution as an upper bound. The global optimum is shown to lie within these two limits.

The solution corresponding to the upper bound is to be used as the results of the solution procedure, as it is declared to be an "approximation" of the actual (global) solution. The model formulation used by Balinski and Mills was the single-plant version of (2.4), that is, a single source, single product, fixed cost warehouse location model. The lower bound they suggested is

$$Z^L = \text{minimum}\left[\sum_j \sum_k C_{jk} X_{jk} + \sum_j \sum_k \left(\frac{f_j(a_j)}{a_j}\right) X_{jk}\right] ,$$

subject to:

Same constraints as (2.3),

where

$$a_j = \text{minimum (Capacity of warehouse j, sum of customer demand)}$$

The lower bound is simply the solution to the original problem when each warehouse is assumed to operate at some large volume (either its capacity or the total flow of goods passing through the system). If the value of the decision variables found in the lower bound solution are substituted back into the original model (that is, a second set of warehouse throughput levels is determined and used to linearize the model) and the problem re-solved, an upper bound on the optimal solution can be found. This upper bound is

$$Z^U = \text{minimum}\left[\sum_j \sum_k C_{jk} X_{jk} + \sum_j \sum_k \left(\frac{f_j(\sum_k X^L_{jk})}{(\sum_k X^L_{jk})}\right) X_{jk}\right] ,$$

subject to:

Same constraints as (2.4),

where

X^L = value of decision variable found in the lower bound solution.

Balinski and Mills prove [11, p. 8] that the cost corresponding to the (global) optimal solution to the original, nonlinear problem (denoted by Z^0) lies between these two bounds. That is,

$$Z^L < Z^0 \leq Z^U$$

The difference between the optimal solution to a problem (Z^0) and its lower bound (Z^L) may be quite large in some situations. This is to say that the Balinski-Mills lower bound may not be a very useful value for many problems. In an example problem posed by Baumol and Wolfe [10], Z^L and Z^0 differ by some 20 per cent. Even though the effectiveness of this bound is questionable, it may be useful in situations where,

1) the cost functions are path independent (as indicated earlier, this means that $f_{ij}(.)$ is dependent only on the volume shipped along transportation path i-j) as in the simple plant location problem, or
2) the linear cost components of the model dominate the nonlinear, or
3) the optimal solution consist of using only a single warehouse (Kuhen and Hamburger point out this latter case, but its validity is not obvious).

The model investigated by Balinski and Mills is a fixed cost type formulation. Their solution procedure is geared to this specific model. However, as in many other approaches, it appears a simple task to extend the solution technique to a general nonlinear model consisting of strictly concave cost. In the above calculations, for example, it makes little difference whether

$$f_j(\cdot) = a_j + b_j X_j,$$

or

$$f_j(\cdot) = b_j X_j^{\frac{1}{2}}.$$

since they both can be converted to a linear expression in X_j (that is, linearized) and the converted problem solved as before.

Another approach reported in the literature which may be classified as a successive linearization technique is that presented by Hammond in 1963 [12]. An excellent discussion of significant and relevant cost factors is given in his paper, but the mathematical model presented is a somewhat simplified version of reality. Essentially, it is the same fixed cost formulation given by Kuehn and Hamburger (2.4). Many of the more complex cost factors discussed by Hammond (such as service time limits) are not explicitly treated in his model. The model does, however, include the possibility of direct shipments from plants to customers without having to go through intermediate storage points. No other location model reported in the literature explicitly incorporates this consideration. The cost of shipments along this direct path are assumed to be linear.

Hammond reports [12, p. 126] that the total cost function for warehouse operation is convex, consisting of fixed and linearly varying portions. Obviously, this type function is actually concave in form. Nevertheless, his approximate solution procedure is not affected by this error. The procedure, flow charted in figure 8, is based on linearizing the nonlinear cost functions in the model in successive steps. At each stage this is done by approximating or

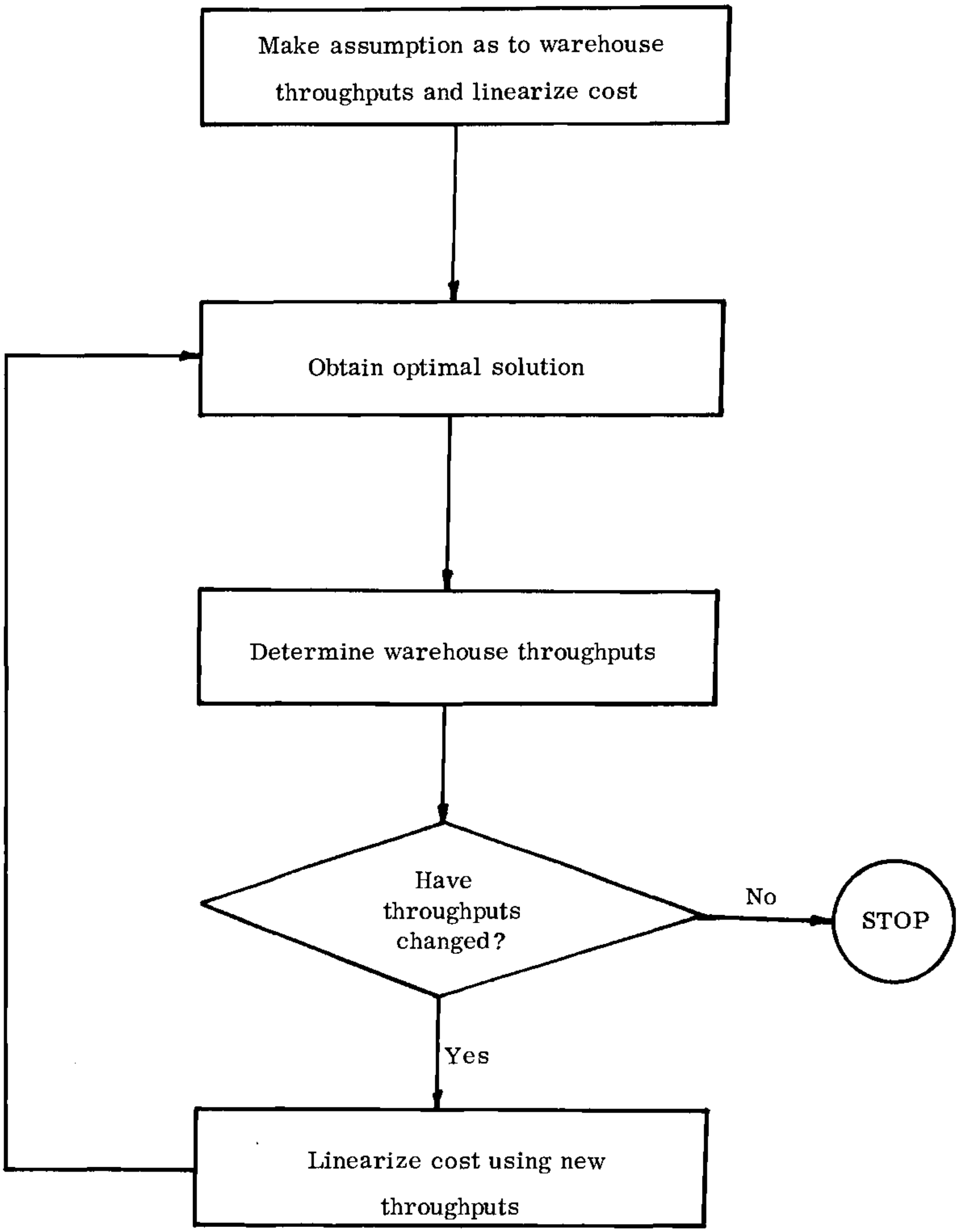


Figure 8. Hammond's Successive Linearization Procedure.

linearizing the total cost function for some assumed throughput level for each warehouse, obtaining unit cost for each warehouse cost function. Mathematically, the total cost function for warehouse j,

$$TC_j = F_j + C_j \cdot \sum_i \sum_k X_{ijk} \quad ,$$

is linearized to yield the unit cost formulation

$$T\bar{C}_j = \sum_i \sum_k X_{ijk} \cdot \frac{F_j}{\sum_{\bar{i}} \sum_{\bar{k}} X_{\bar{i}j\bar{k}}} + C_j \quad ,$$

where

F_j = fixed cost of operating warehouse j,

C_j = variable cost of operating warehouse j,

which is added to the transportation cost for each path involving warehouse j in the location model.

After each linearization, the linear model is solved and the solution used for the next linearization. As the flow chart indicates, Hammond continues this cycle until two successive throughput sets (a throughput set being the collection of throughput levels for all warehouses) are equivalent. This is the same stopping rule used by Baumol and Wolfe. However, for large problems it is not difficult to envision different solutions (assignments of customer demand to specific warehouses or plants) resulting in identical throughput sets. Since the total system's cost would probably vary with different solutions, a more logical decision rule for stopping the algorithm would be to halt after two successive solutions (customer assignments) are the same.

An important difficulty inherent in a location-allocation problem not treated by any of the successive linearization approaches reported in the literature is the consideration of the effects of combining individual operating cost curves in the objective function of the model to be solved. That is, when there are two or more possible warehouses to assign a customer to and the total cost functions for two or more of these alternative assignments intersect, special solution techniques are needed to avoid suboptimal results. An example problem illustrating this inadequacy specifically in Hammond's algorithm is given in the appendix. Graphically, this difficulty is seen in figure 9.

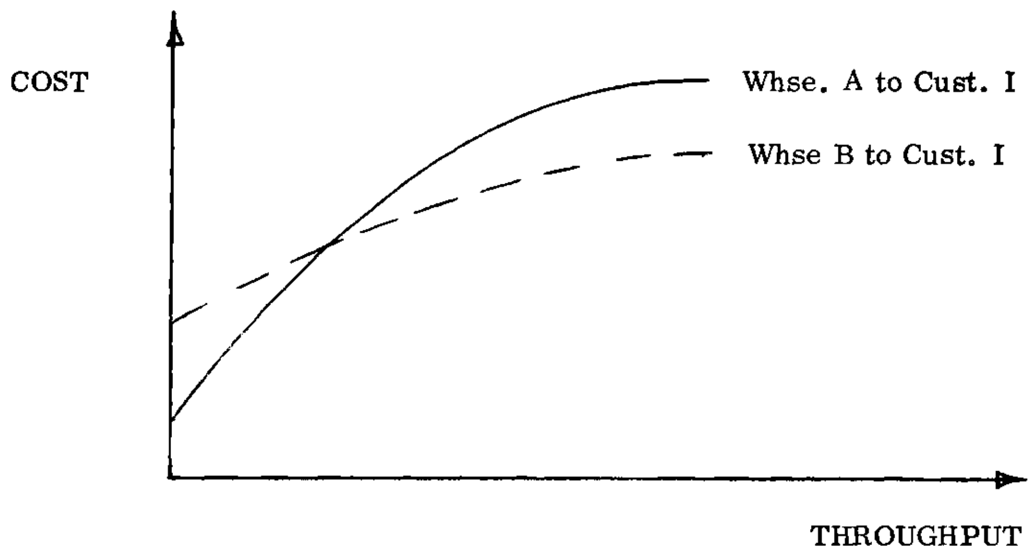


Figure 9. Intersecting of Operating Curves.

Recent Efforts

The majority of the current literature concerning location-allocation problems is devoted to mixed integer formulations. In particular, effort is seemingly concentrated on improving the efficiency of the imbedden linear programs, or on improving branching and/or bounding decision rules. Sa'\'s research [5], along with White's [13], and Spidelburg's [14], is an example of this direction in the recent location literature.

Summary

Distribution planning is important both as a corporate-level concern and as a theoretical problem. Through proper formulation and solution of warehouse location models, investment planners can save the firm significant amounts of capital expenditure. However, the basic structure of the location problem causes solution difficulties that have commanded a great deal of attention in the literature. The reason for these difficulties stems from the fact that the location model calls for the minimization of a concave function over a convex set of constraints.

Corresponding to various formulations of this concave cost function, several different solution approaches have been advocated in the literature. These approaches can be grouped into four different categories: mathematical programming, simulation, heuristics, and successive linearization. Solution techniques for the pure warehouse location problem are actually a subset of approaches for the more general location-allocation problem, with the wealth of relative literature pertaining to this latter area.

Each of the categories of solution approaches has its advantages and disadvantages. Mathematical programming, for example, produces a result which is a global optimum when the model is cast in the mixed-integer mode. However, this approach has the disadvantage that large, practical problems may require too much storage and computational resource.

On the other hand heuristic and successive linearization methods can effectively cope with large problems having many decision variables. Also, more realistic modeling is possible with these approaches than with the mixed-integer approach. However, these techniques cannot guarantee to produce global or best solutions and may result in a solution which is optimum merely in a region of values for the decision variables. The linearization technique also offers the disadvantage of having an inherent bias in its solution toward possessing too many facilities, causing high-cost solutions.

Simulation models can be formulated with a high degree of detail, including stochastic considerations. However, computer resource requirements are high since each extreme point (serving as a decision alternative) must be examined to ensure global results. Total enumeration of this type is impractical for large problems. However, if a decision strategy could be developed for moving efficiently from one alternative solution to another, simulation might become a very competitive solution approach for location problems.

CHAPTER III

MODEL FORMULATION

The model developed in this research is a representation of the physical distribution system common to many firms having a nation-wide customer market serviced by common carriers. The system consists of multiple production facilities (with fixed locations and capacities) manufacturing many individual products that may be delivered to any of several thousand customers. Intermediate storage may be utilized as public or leased warehousing space. Any of several available common carrier modes of transportation may be utilized along any supply or resupply path (a path being a source to sink route employing a particular transportation mode). Goods may be shipped to customers "direct" from the factory or from an inventory at an intermediate warehouse. Also included in the distribution system is a delivery time constraint which must be met by each supply path utilized. (A supply path is a path from a plant or warehouse to a customer sink; a resupply path is a path from a plant to a warehouse).

Distribution System

The system can be further explained by tracing the flow of finished goods through the various system components. All production and distribution functions associated with the products before they are converted into finished goods are considered fixed. This includes production capacities, production schedules, and plant locations. These functions are considered as constant system components and will not be subject to analysis in the distribution model.

Plants

After being produced a product has two alternate routes it can take. It may be directed to an in-house storage, or staging, area for short or long term holding, from which it will be shipped "direct" to a customer. This storage area may be part of the physical plant itself, or an external storage facility in the proximity of the plant. In either case, for purpose of analysis, the storage area may be appropriately thought of as a warehouse located at the plant site. For purposes of costs analysis, plant staging may be thought of as adding a unit warehousing cost to each plant-to-customer path.

Alternatively, a finished good may bypass the staging area and be shipped to an outlying distribution-type warehouse. In this case, finished goods move directly from the assembly line to transportation carriers and then on to a warehouse. This warehouse may be either a public or a leased warehouse.

Warehousing

Public warehousing involves paying an external source for all handling and storage functions involved in warehousing goods between the time they reach the warehouse and the time they are transported away from the warehouse. On the other hand, when warehousing is of the leased type, the building or space used is leased from an external source while the handling and storage functions are performed by company-owned resources. Normal operating expenses such as utilities are considered part of the leasing agreement and hence part of the leasing fee.

All products stored in a warehouse are considered to be "shelf" type goods. That is, they are subject to reordering and storage according to a formal inventory policy. In the current model, a simple "EOQ" policy* is assumed to be used for all warehouses regardless of type.

Transportation

Several alternative transportation modes are available for transporting final goods along most supply and resupply paths. Examples of these are truck, rail, and piggyback. All modes are assumed to be common carriers. Some paths may not have all modes available since some cities are not accessible or not serviced by some carriers. The capacity of each mode of carrying goods along any serviced paths is assumed to be unlimited.

The supplying of a final good to a customer must be performed within a certain time limit after that customer's order has been received. This implies that only those paths associated with sources capable of delivering goods to customer within this service limit are feasible supply paths. For example, if a supply route were Atlanta, Ga. (source) to Seattle, Wash., (sink) and delivery of goods by rail took more than a service limit of, say, three days, then the Atlanta-Seattle-rail path is an infeasible alternative for supplying final goods to Seattle.

Customers

Final goods shipped through a nation-wide distribution system such as the one under present consideration typically are demanded by several thousand

*A simple "EOQ" inventory policy refers to the fixed-quantity reorder lots and reorder points associated with a total cost model involving only holding and ordering cost. No storages are allowed.

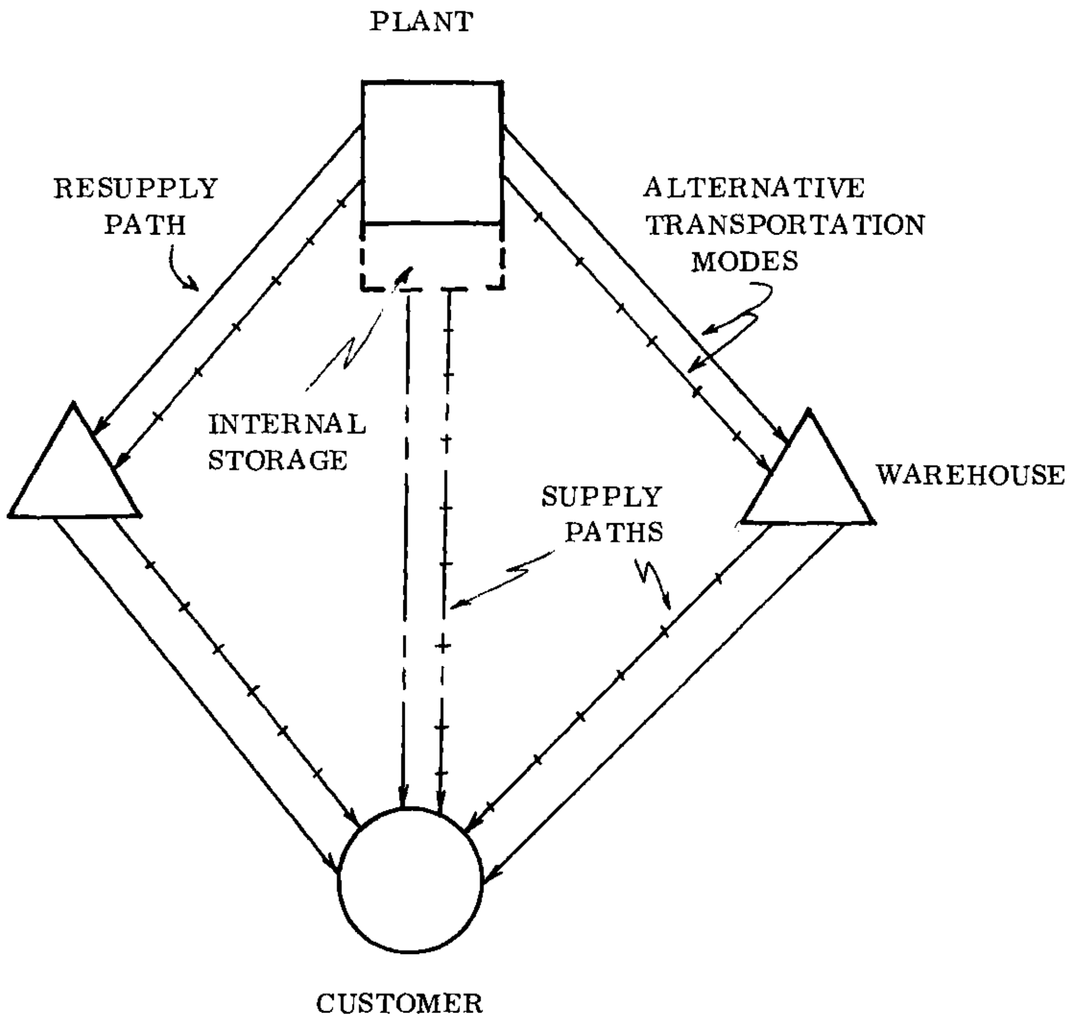


Figure 10. Diagram of the Physical Distribution System

(Note that this is a single plant, single customer case).

individual customers. Each customer may have a demand for any of several different products manufactured by the firm. This demand, assumed to be deterministic in nature and based on an annual period, may exhibit seasonal or trend patterns. Associated with each customer's demand for each product is an average order size. This is simply the average size (averaged over the annual time period involved) of shipments of each product sent to each customer from all sources. The purpose of average order size determination is to serve as the basis in computing transportation costs rates. In cases where shipments of one or more products are combined, the appropriate shipment size is the entire shipment since this is the quantity on which transportation costs are based.

A graphic summary of the distribution system is given in figure 10. This diagram illustrates the relationships of the basic system components. Once these components and their relationships have been quantified and expressed as a model, the model can be solved and the optimum value of the parameters of interest can be determined. These parameters are the flows of each product along all supply and resupply paths which minimize the total cost of the system and meet all constraints imposed.

General Model

A general formulation of the model of the distribution system previously described is,

minimize: Total System Costs, (3.1)

subject to: (1) All customer demand must be satisfied,

(2) Deliveries must be made within a specified time limit,

(3) Capacities at all production facilities must not be exceeded.

A specific form of the model can be developed by describing 1) the "type" of model, 2) the cost components of the model, 3) the "customer zones" derived, 4) the formulation of "product multipliers," 5) the constraints involved, and 6) the assumptions underlying the model.

Type of Model

Location models can be grouped into two categories or types, those with a finite set of alternative location sites (sites for potential warehouses) and those with an infinite set of alternative sites. There are advantages to employing either type. For example, an obvious advantage of the infinite set model is that the solution will not fail to utilize an otherwise profitable site simply because it was not included in the list. However, a definite disadvantage of this type model is that the model must be formulated in terms of the straight-line distance formula-- $((\bar{X}-X)^2 + (\bar{Y}-Y)^2)^{\frac{1}{2}}$. The nonlinearities introduced due to this formula are extremely difficult to handle. In addition, use of this distance formula forces all transportation costs used in this type model to be based on approximate mileage cost rates. This need not be the case with a model based on a finite set of sites, since actual cost rates can be used. Also, a finite list type model has the advantage that the need for the nonlinear distance formula is eliminated.

The advantages of both modeling approaches (or types) may be incorporated into the present model by formulating what might be referred to as "location zones." A location zone can be defined as a geographical area within which one and only one warehouse may be located. In other words, a location zone is a pseudo warehouse location site. The entire continental United States can be broken up into location zones so that every possible location site is contained within some zone. In this manner, an infinite number of potential sites are being implicitly considered while explicit consideration is being given to the finite list of location zones.

If it were decided (through solving the model) that a particular zone should have a warehouse located within it, a secondary investigation would be made to decide exactly where to locate the warehouse within the zone. This secondary decision should be based not only on quantitative considerations but also on subjective factors. For example, consideration should be given to labor availability, customer relations, and relative location of competitors within the zone.

The size and shape of zones should vary according to several factors. In areas of little or no customer demand there is no necessity in having a large number of alternative location sites. Obviously, a solution analysis is not likely to select areas of sparse customer demand as location points for warehouses. There are more profitable, or least-cost, sites--namely, areas of high demand density. Therefore, location zones in areas where demand is of low intensity may be larger and take in more geographical area than zones in high intensity areas. That is, there need be fewer location site alternatives--implying fewer, larger zones--in regions of low customer density, and more site alternatives--implying more, smaller zones--in high density regions.

Geographical considerations also should play a part in dictating the relative size of zones. For obvious reasons, there need not be a zone encompassing the Smoky Mountains, or the Great Lakes. Only one alternative location site need be used to consider most of the dessert lands of the West.

State boundaries are significant factors in determining the specific shape of zones. Since tariff rates and transportation costs are significantly affected by the crossing of state boundaries, zones defined so that such crossing is kept to a minimum will encompass more cost-homogeneous areas.

A final consideration in zone defining is that the total number of zones should be kept to a minimum, implying inflation of zone sizes. Since tariff rates for all combinations of plant to warehouse (or location zone) and warehouse to customer paths will have to be obtained for solution of the model, keeping the number of zones to a minimum will ensure a more reasonable amount of required data gathering. Balanced against this desirable goal or reduced data requirements is the objective of giving explicit consideration to as many alternative sites as possible (implying smaller zone sizes). This trade-off of objectives must be dealt with as a significant subjective factor in defining zones.

In order to evaluate the cost of shipments from one zone to another, some type of geographic reference point (that is, a city) must be designated within each zone. Based on these reference points, or "key cities," point-to-point transportation cost data can be gathered. The primary consideration which should be used in designating a particular city within a zone as that zone's key city is the probability that the key city would be chosen (in the secondary analysis mentioned previously) as the actual warehouse location site within that zone. Other factors which influence this selection and which must be given consideration are 1) that the availability of transportation service for the key city be representative of service for the entire zone, and 2) the tendency of the key city to be the geographical center of the zone.

To reiterate, the distribution model is based on a finite list of warehouse location zones. Each zone is a pseudo warehouse location site. The use of such a "location zone" concept has several advantages. Among these are:

(1) all possible locations sites are either explicitly or implicitly considered by the model,

(2) the model avoids the solution difficulties inherent in an infinite site model,

(3) cost data can consist of actual point-to-point cost of shipments, rather than being based on approximate cost rates,

(4) qualitative location factors can be considered and used as well as quantitative ones.

Throughout the remainder of this paper the term "warehouse" will be used to refer to either a location zone or to the physical warehousing facility itself.

Model Cost Components

There are two basic components which make up the total cost function of the model (3.1). These are transportation costs and warehouse operating costs. A brief description of the characteristics of each component will be given.

Transportation Costs. There are two types of transportation paths represented by the model, supply and resupply paths. Recall that a path was defined as the transportation route between a source (either a plant or a warehouse) and a sink (either a warehouse or a customer) utilizing a particular transportation mode. The tariff rates, or transportation cost rates, associated with supply paths (that is, any path with a customer as a sink) decrease with increasing intervals of volume transported. There is assumed to be no fixed cost associated with the use of any supply path. Since the amount or volume transported along a supply path in any one shipment is, on the average, equal to that customer's expected or average shipment size, the unit transportation cost for that path is a function of that average shipment size. Once each customer's average shipment size has been determined, unit cost rates for each supply path can be found.

Resupply paths possess tariff rate structures similar to those associated with supply paths. However, unlike supply paths, unit transportation costs along resupply paths are not dependent on customer shipment size. They are a function of inventory reorder quantities requested by warehouses. Normally, the service time or lead time associated with delivering this resupply quantity is not critical and can be extended or shortened by decisions at the plant site without seriously affecting warehouse operations. This may allow shipment quantities to build up to sizes large enough to take advantage of the least-cost tariff rates available. That is, a resupply plant may decide to consolidate resupply shipments to a particular warehouse in order to capitalize on a more economical tariff rate.

Also, normal reorder quantities may be of a magnitude large enough to qualify them for least-cost rates.

Obviously, the feasibility of making use of such resupply policies depends upon the products and processes in the actual distribution system being represented. For the current model, the assumption will be made that reorder quantities requested by warehouses can be manipulated so as to take advantage of the least-cost tariff rate for each transportation mode available.

Most common carriers exhibit very similar tariff schedules. That is, tariff rates for each mode decrease in discrete jumps as the tonnage shipped increases. An example of such a rate structure is that for rail shipments from Tampa, Fla. to Dothan, Ala., which might be:

Volume (100 lbs.)	<50	50-70	> 70
Tariff Rate ($per 100 lbs.)	.96	.85	.77

It is assumed that all alternative modes utilized in the model exhibit this type rate structure.

A final aspect of transportation cost which should be mentioned is that the cost associated with the flow of goods along transportation paths may be more appropriately represented by transportation cost plus unit production cost rather than transportation cost alone. This allows more weight in the solution analysis to be given to those sources associated with lower production cost. All other factors being equal, the model should allow the most economical production facility to produce all goods. Therefore, shipping costs (hereafter called transportation cost) along all resupply and plant-to-customer supply paths in the model should include production costs. Costs associated with all other paths should include only pure transportation cost. Otherwise production cost would be counted twice in plant-to-warehouse-to-customer paths.

Warehouse Operating Costs. Each of the two alternative types of external warehousing available in the model, leased and public warehousing, has a distinct operating cost function. Both, however, have the same major cost components--handling and storage costs. As explained earlier, public warehousing involves paying external sources for all handling and storage costs. One such payment consist of an in-out or throughput charge to cover handling of all goods flowing through this type warehouse. That is, each item passing through a public warehouse incurs a unit handling charge. There is also a monthly storage charge for all items in inventory at the beginning of each month. That is, all goods in stock at the beginning of a month are assumed to remain in stock the entire month, even though half the inventory may be shipped out on the second day of the month. Therefore, storage charges are incurred by those goods actually in the warehouse on the first of the month.

Since a fixed-quantity reorder policy is being assumed by the model, the beginning of the month inventory level is simply the average inventory level at each warehouse. That is, no attempt is being made by the appropriate decision makers in the physical distribution system to minimize this beginning inventory level. Due to the complexities of the charging system of public warehouses (some of which are not included in the model), and to the somewhat unreliable lead times of most common carriers, little effort is actually devoted in practice

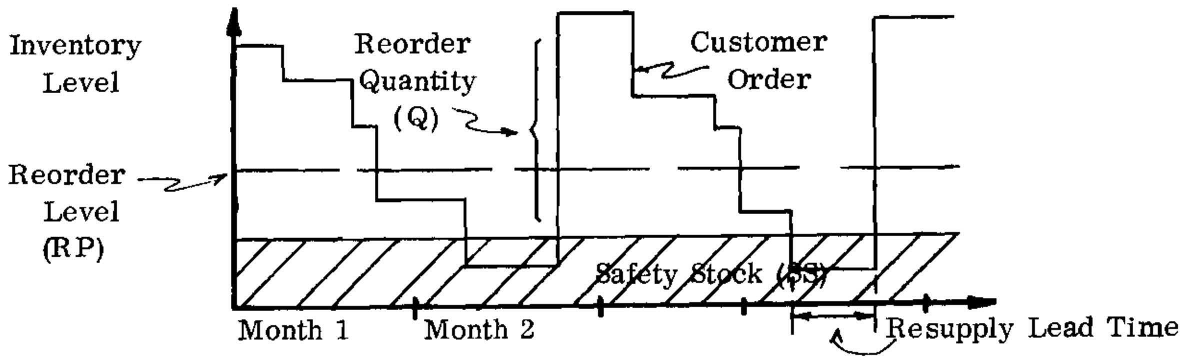


Figure 11. Warehouse Inventory Policy.

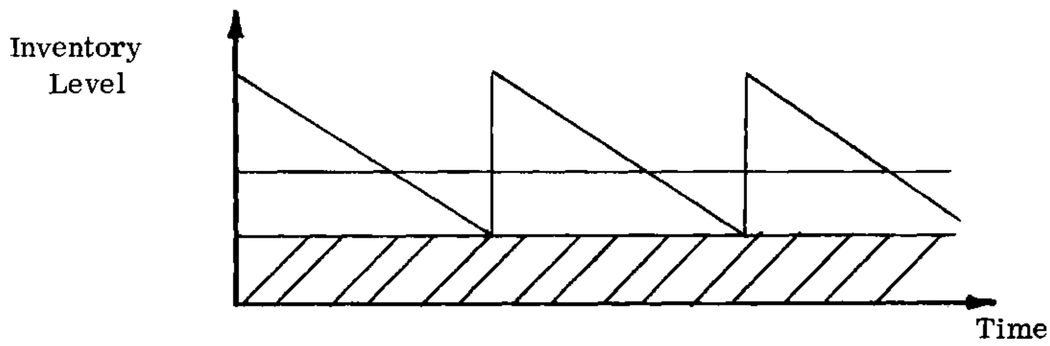


Figure 12. Average Inventory Level ($\frac{1}{2}$Q+SS).

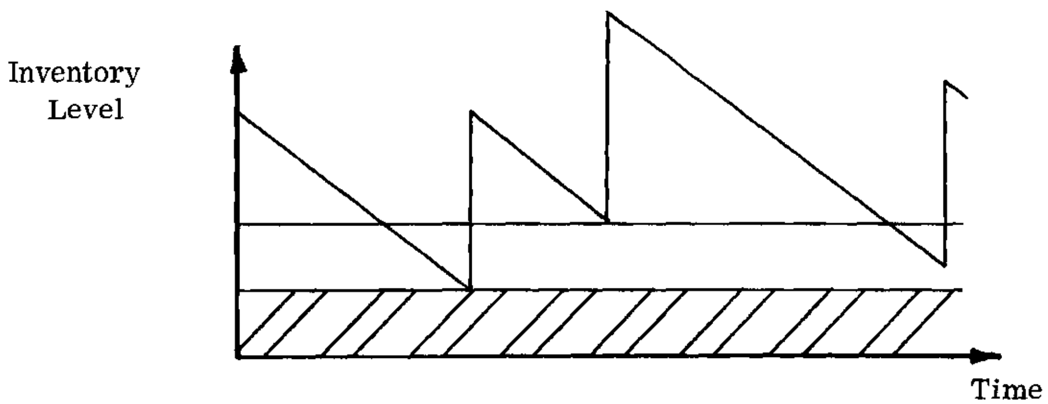


Figure 13. Maximum Inventory Level (Q+RP).

to reducing first of the month inventory levels.

The average inventory level at each warehouse j (under the simple EOQ policy of the model) is

$$I_{AVE} = \sum_m \left[\tfrac{1}{2} \cdot Q_{jm} + SS^{w}_{ijm}\right], \tag{3.2}$$

where

Q_{jm} = reorder quantity at warehouse j for product m,

SS^{w}_{ijm} = the amount of safety stock of product m carried by warehouse j when resupplied by plant i using transportation mode w.

Figure 11 through 13 graphically develop this average inventory level for the single warehouse, single plant, single product case.

Since Q_{jm} in (3.2) is equal to a constant multiplied by the square root of the annual demand (or throughput) satisfied by warehouse j for product m, the average inventory level can be rewritten as

$$I_{AVE} = \sum_m \left[\tfrac{1}{2}C_{jm} \cdot (D_{jm})^{\frac{1}{2}} + SS^{w}_{ijm}\right], \tag{3.3}$$

where $C_{jm} = \left(\frac{2 \cdot \text{order cost}}{\text{holding cost}}\right)^{\frac{1}{2}} = \left(\frac{2C^{o}_{j}}{C_m}\right)^{\frac{1}{2}}$,

D_{jm} = annual throughput at warehouse j for product m.

Using this formulation of the average inventory level, annual storage costs for public warehouse j resupplied by plant i can be expressed as

$$C_{STORE} = 12 \cdot R^{av}_{j} \cdot \left[\sum_m \left[\tfrac{1}{2} \cdot \left(\frac{2C^{o}_{j}}{C_m}\right)^{\frac{1}{2}} \cdot (D_{jm})^{\frac{1}{2}} + SS^{w}_{ijm}\right]\right], \tag{3.4}$$

where R_j^{av} = monthly storage charge per unit at warehouse j.

Since handling cost is a linear function of throughput, this cost component may be expressed as

$$C_{HANDLE} = R_j^{in} \cdot \left[\sum_m D_{jm} \right] , \tag{3.5}$$

where R_j^{in} = unit handling charge at warehouse j.

Combining these two components, annual warehouse operating cost for public type warehouse j can be expressed as

$$TC_{PUBLIC} = R_j^{in} \cdot \left[\sum_m D_{jm} \right] + 12 \cdot R_j^{av} \cdot \left[\sum_m \left[\tfrac{1}{2} \cdot \left(\frac{2C_j^o}{C_m} \right)^{\frac{1}{2}} \cdot (D_{jm})^{\frac{1}{2}} + SS_{ijm}^w \right] \right] . \tag{3.6}$$

Complicating this expression is the fact that R_j^{in} and R_j^{av} are not constant rates but rather are a function of the warehouse throughput, $\sum_m D_{jm}$. This occurs because handling and storage rates are subject to renegotiation. That is, public rates can normally be reduced at specific throughput volume break-points due to favorable reaction of the public warehouse to an increase in business. This is analogous to the decreasing rate structure of common carriers discussed earlier. For purposes of analysis, however, these rates will be expressed as though they are constant. Their variability will be discussed in connection with the solution of the model.

As previously indicated, leased warehousing refers to the use of company-owned resources to operate leased warehouse space. There is a fixed cost associated with these company-owned resources. Normally, these resources are a lift truck, a lift truck operator, and a shipping clerk. Administrative supplies are assumed to be insignificant. A leasing fee is paid to the owner of the physical warehouse for use of the building, including utilities, insurance, and the like.

The amount of space leased must be sufficient to house the maximum amount of goods expected to be in the warehouse at any one time. Consequently, leasing charges are a function of the maximum inventory level which may be experienced by a warehouse.

As indicated in figure 13, under the model's EOQ policy the maximum inventory level that a warehouse (j) might experience is (note that this formulation will result in leasing space which is, to a large extent, unutilized)

$$I_{MAX} = \sum_m \left[Q_{jm} + RP^{w}_{ijm} \right], \tag{3.7}$$

where RP^{w}_{ijm} = the inventory level used at warehouse j as a reorder point for product m when j is resupplied by plant i via mode w.

Replacing Q_{jm} by its equivalence, $\left(\frac{2C^{o}_{j}}{C_m}\right)^{\frac{1}{2}} (D_{jm})^{\frac{1}{2}}$, the expression for the maximum inventory level can be rewritten as

$$I_{MAX} = \sum_m \left[\left(\frac{2C^{o}_{j}}{C_m}\right)^{\frac{1}{2}} \cdot (D_{jm})^{\frac{1}{2}} + RP^{w}_{ijm} \right]. \tag{3.8}$$

Letting R^{L}_{j} represent the annual charge per unit held in storage (that is, the per-unit leasing fee), leasing cost at leased warehouse j--resupplied by i--can be written as

$$C_{LEASE} = R_j^L \cdot \left[\sum_m \left[\left(\frac{2C_j^o}{C_m} \right)^{\frac{1}{2}} \cdot (D_{jm})^{\frac{1}{2}} + RP_{ijm}^w \right] \right] . \tag{3.9}$$

Adding on the fixed cost associated with inventory handling and storage costs, the expression for total operating cost in a leased warehouse (j) is

$$TC_{LEASED} = FC_j + R_j^L \cdot \left[\sum_m \left[\left(\frac{2C_j^o}{C_m} \right)^{\frac{1}{2}} \cdot (D_{jm})^{\frac{1}{2}} + RP_{ijm}^w \right] \right] . \tag{3.10}$$

In both leased and public warehousing, operating costs are dependent upon the particular resupply path used to restock warehouse j's inventory. This is due to the dependence of both SS_{ijm}^w and RP_{ijm}^w on the source (i) and the transportation mode (w) used. If the lead time associated with warehouse j's resupply path (there will normally be only one such path) is relatively short, then the safety stock and reorder point levels will be low. To see this, recall that the purpose of carrying safety stock is to provide a "buffer" stock to cover unpredictable demands during the resupply lead time. If this lead time is of a long duration, a greater amount of buffer stock is needed as the required insurance against shortages due to the uncertain demand process. If the lead time is short, less buffer is required as there is less of an opportunity for an unpredicted demand to occur which might result in a shortage. To take an extreme example, if a warehouse had instantaneous resupply (no lead time), there would be no chance of a shortage, hence no need for a safety stock at all.

Similarly, the reorder point is directly related to lead time duration. Reorder point can be defined as the maximum "reasonable" demand during lead time. The shorter the lead time, the smaller the maximum demand during lead time. If resupply were instantaneous, the reorder point would be a zero inventory level. Hence, it is obviously desirable to minimize the lead time for resupplying each warehouse; that is, to utilize resupply paths that have the shortest lead time.

However, reducing the resupply lead time may be costly. As mentioned earlier, a resupply path's lead time is a function of the path's source and its transportation mode. Given that a warehouse will be supplied from the closest plant (provided capacities are not exceeded), then reducing the lead time entails choosing the fastest transportation mode. But faster modes are normally more expensive. For example, it takes approximately four days to ship 3600 pounds from Atlanta, Ga., to Dallas, Texas, by truck. This cost approximately $1.12 per 100 pounds. On the other hand, shipping 3600 pounds from Atlanta to Dallas by rail takes 14 days but costs only $.96 per 100 pounds.

Therefore, there is an inherent trade-off in the distribution system between inventory and transportation costs. Reducing inventory costs can result from reducing the resupply lead time. But reducing the lead time results in using a more expensive means of transportation. Likewise, reduced transportation costs can result from using slower, less expensive modes. The use of slower modes, however, means that safety stock and reorder levels will have to be larger to cover uncertainties in demand during this enlarged lead time. Consequently, inventory costs will increase. This cost trade-off will be explicitly handled in the solution procedure of the model.

Customer Zones

Most practical, large-scale distribution systems entail deliveries to several thousand individual customers. Goods may be shipped straight from production facilities or they may go to regional warehouses before being dispersed to the final customer. Since the current model is based on describing and analyzing each individual supply and resupply path, the number of such paths is of utmost concern as regards the efficiency of the solution of the model. Also, practical models should stress minimum data-gathering requirements. Obviously, data requirements for evaluation of, say, 2000 (customers) multiplied by 100 (sources) equaling 200, 000 supply paths would be enormous.

By forming "customer zones," the number of final destinations, and hence supply paths, can be kept to a reasonable number. A customer zone, analogous to a warehouse location zone, can be defined as a geographical area encompassing many individual final customers. For example, the state of Georgia might serve as a customer zone and have several hundred individual customers located within it. Or, the city of Atlanta might serve as a zone by itself. A customer zone is intended to represent the demand and ordering patterns of the individual customers within the zone. In this manner, zones substitute for individual customers in the model.

The guidelines for determining the boundaries of customer zones are similar to those used in defining the warehouse location zones discussed earlier. Specifically, the size of each zone should be inversely proportional to the customer density in the area of the zone. High density regions should have more, smaller customer zones than low density regions. For example, if product demand is strongly correlated with population, densely populated areas should have many smaller zones, allowing each zone to better represent the ordering patterns and habits of the individual customers within that area.

Unlike location zones, customer zones do not have to be contiguous. In large areas of no customers, there is no need for a zone at all. However, where zones are formed their shapes should be influenced by state boundaries and geography as are the shapes of location zones. There should be no gerrymandering of zones.

An obviously desirable feature of a zone is to contain individual customers having similar ordering and demand characteristics. The use of this feature should be given heavy weight in the formation of zones since a zone is to be used as a pseudo-customer, representing the individual customers within it. That is, each customer zone acts as a sink in the distribution model. Each zone has an associated annual demand (this is the sum of the annual demands of each individual customer within the zone).

As do individual customers, zones have an average or expected shipment size for all products for the year which the model is to represent. All transportation costs to a zone are based on these average shipment sizes, which represent the average size of shipments (for each product) to all customers within a zone. That is, they should be composite averages. Where order sizes vary greatly from one customer to another within zones, the expected shipment sizes should be some type of weighted averages. Due to the fact that tariff rates at a decreasing rate as the shipment sizes increase, larger shipments should be given more weight in the computations. However, the assumption will be made for the current model that ordering characteristics (order size and frequency) do not significantly differ for customers within relatively close proximity of one another. This allows the expected shipment size of a zone to correctly represent the average size of shipments to all customers within the zone without applying weighting factors.

Associated with each customer zone must be a reference point, or a "key city." Such a point is necessary to evaluate path transportation costs for all paths involving that zone. The key city for a zone should be the point which minimizes the average distance (or the expected tariff rate based on each customer's expected order size) from the point to each customer within the zone, the distance being weighted by the customer's annual demand. In this manner, supply path transportation costs, calculated using the key city as the sink in a point-to-point rate determination, will closely approximate cost of actual shipments to individual customers within the zone. That is, transportation cost of a shipment from a sink to the key city of a zone will equal the average cost of that shipment to all customers within the zone.

Other influences in the designation of key cities in customer zones are the following two factors.

(1) Cities in areas where the density of competitors is low should be given more weight than areas of high competition density. Demand around points laden with competitors is not likely to be a stable demand, representative of the demand structure in the entire zone. Even though it may be the demand center of the zone today, all customers in a high density area may be lost tomorrow.

(2) Cities representing the availability and cost structure of transportation throughout the zone should be heavily weighted.

Product Multiplier

The model used to represent the distribution system assumed that "product" demand refers to individual products or items. However, as was the case with the large number of individual customers, the number of distinct items being distributed in the system is normally too large to allow explicit consideration of each item's flow in the model. This necessitates consolidating individual items into product groups or lines. However, use of data based on product lines rather than products creates an understatement of the costs involved in the model. Specially, warehouse operating costs are less when product lines are used than when the formulation explicitly considers individual products.

This occurs because the "approximate" model (using product lines) bases inventory costs on the square root of the sum of warehouse throughput of all items (for each warehouse), whereas the true model (using individual items) would base inventory costs on the sum of the square roots of the throughput for each item (for each warehouse). In other words, the square root of a sum is being used to estimate the sum of the square roots of the components of the sum.

In the model, actual inventory costs for warehouse j resupplied by plant i via mode w should be represented as

$$\text{a)}\quad IC^{P}_{ACT} = R^{av}_{j} \cdot \left[\sum_{m=1}^{NM} \left[\sum_{\bar{m}=1}^{(N\bar{M})_m} \left[\tfrac{1}{2} \cdot C_{jm} \cdot (D_{jm\bar{m}})^{\frac{1}{2}} + SS^{w}_{ijm\bar{m}} \right] \right] \right] \qquad (3.11)$$

$\{$Public warehousing

and

b) $$IC^L_{ACT} = R^L_j \cdot \left[\sum_{m=1}^{NM} \left[\sum_{\bar{m}=1}^{(N\bar{M})_m} \left[C_{jm} \cdot (D_{jm\bar{m}})^{\frac{1}{2}} + RP^W_{ijm\bar{m}}\right]\right]\right] \quad \Big\{ \text{Leased warehousing}$$

where $D_{jm\bar{m}}$ = the throughput or annual demand of item $\bar{m}$ in product line m at warehouse j,

$C_{jm}, SS^W_{ijm\bar{m}}, RP^W_{ijm\bar{m}}$ = the inventory cost constant, safety stock and reorder point coefficients, respectively, described earlier with the item subscript, $\bar{m}$, added,

NM = the upper limit of m,

$(N\bar{M})_m$ = the number of items in product line m.

When items are grouped into product lines and the model based only on these lines, inventory costs are approximated by

a) $$IC^P_{APP} = R^{av}_j \cdot \left[\sum_{m=1}^{NM} \tfrac{1}{2} \cdot C_{jm} \cdot \left(\sum_{\bar{m}=1}^{(N\bar{M})_m} D_{jm\bar{m}}\right)^{\frac{1}{2}} + \sum_{\bar{m}=1}^{(N\bar{M})_m} SS^W_{ijm\bar{m}}\right], \tag{3.12}$$

and

b) $$IC^L_{APP} = R^L_j \cdot \left[\sum_{m=1}^{NM} C_{jm} \cdot \left(\sum_{\bar{m}=1}^{(N\bar{M})_m} D_{jm\bar{m}}\right)^{\frac{1}{2}} + \sum_{\bar{m}=1}^{(N\bar{M})_m} RP^W_{ijm\bar{m}}\right].$$

Approximating inventory costs in this manner implies that the following relationship holds:

$$IC^P_{ACT} = IC^P_{APP}, \tag{3.13}$$

or, $$R^{av}_j \cdot \left[\sum_{m=1}^{NM} \left[\sum_{\bar{m}=1}^{(N\bar{M})_m} \tfrac{1}{2} \cdot C_{jm} \cdot (D_{jm\bar{m}})^{\frac{1}{2}}\right]\right] = R^{av}_j \cdot \left[\sum_{m=1}^{NM} \tfrac{1}{2} \cdot C_{jm} \cdot \left(\sum_{\bar{m}=1}^{(N\bar{M})_m} D_{jm\bar{m}}\right)^{\frac{1}{2}}\right],$$

and a similar relationship involving R^L_j terms. This is saying that the sum of the square root of a series of components is equal to the square root of the sum of the components. In simplier terms, the mathematical relationship that is implied is

$$\sum_{i=1}^{N} (A_i)^{\frac{1}{2}} = (\sum_{i=1}^{N} A_i)^{\frac{1}{2}}, \tag{3.14}$$

where $A_i \geqslant D_{jm\bar{m}}$.

Clearly, this relationship only holds when either a) $A_i = 0$ for all i, or b) only one A_i is positive. This latter case can be interpreted as saying that only one item within a product line has a positive throughput. Extended to the general model, this implies that each product line (m) at each warehouse has only one item within it with a positive warehouse throughput. Obviously, when product lines and items coincide in the distribution system (that is, product lines are not utilized) this equivalence is implied.

When this is not the case, that is, when there is more than one positive throughput in each sum and hence the sum of the square roots is less than the square root of the sum, the $(\sum_{i=1}^{N} A_i)$ terms in the model must be multiplied by some factor. Use of the proper multipliers should force the equivalence to hold between actual inventory costs and inventory costs as represented by the square root terms of the model. For example,

$$IC^P_{ACT} = N_m \cdot IC^P_{APP}, \tag{3.15}$$

or,

$$R_j^{av} \cdot \left[\sum_{m=1}^{NM} \left[\sum_{\bar{m}=1}^{(N\bar{M})_m} \tfrac{1}{2} \cdot C_{jm} \cdot (D_{jm\bar{m}})^{\frac{1}{2}} \right] \right] = R_j^{av} \cdot \left[\sum_{m=1}^{NM} N_m \cdot (\tfrac{1}{2}) \cdot C_{jm} \cdot \left(\sum_{\bar{m}=1}^{(N\bar{M})_m} D_{jm\bar{m}} \right)^{\frac{1}{2}} \right]$$

where N_m = the multiplier for the m^{th} product line; that is, the multiplier for the m^{th} sum.

Since the throughput of each item at all warehouses cannot be predicted in advance, one logical approach to formulating the required multipliers is as follows: determine the value and form of each multiplier when the maximum deviation of approximated costs from true costs occurs, and the value when the minimum deviation occurs. One of these two values (or one in between) can be utilized as a multiplier in the solution process. Then, a sensitivity analysis can be performed in order to determine the influence of changes in the multipliers. If it is established that changes in the multiplier values significantly affect the problem solution, an external investigation should be performed to determine the actual throughput of each item. These throughputs are derived from the warehouse assignments of each customer-sink. From the knowledge of each customer's demand for each item, the actual multiplier can be established which equates approximate and true inventory costs at each warehouse. This is the approach employed in the current model and solution procedure.

As pointed out previously, the minimum deviation of approximate from actual costs occurs when no more than one of the sum components are positive. This results in the equivalence (again in simplified notation)

$$\sum_{i=1}^{N} (A_i)^{\frac{1}{2}} = (\sum_{i=1}^{N} A_i)^{\frac{1}{2}}. \qquad (3.16)$$

The appropriate value of the multiplier, N_m, is one. That is,

$$\sum_{i=1}^{N} (A_i)^{\frac{1}{2}} = N_m \cdot (\sum_{i=1}^{N} A_i)^{\frac{1}{2}}, \tag{3.17}$$

where $N_m = 1$.

On the other hand, the maximum deviation of costs occurs when the throughputs at a warehouse of all items are the same (within a product line). That is, when $D_{jm\bar{m}}$ is the same for all $\bar{m}$, the sum of the components is equally divided among the components and deviation of costs is larger than that corresponding to any other dispersion of the sum. To see that this is so, note that the deviation can be expressed as

$$\sum_{i=1}^{N} g(A_i) = \sum_{i=1}^{N} (A_i)^{\frac{1}{2}} - (\sum_{i=1}^{N} A_i)^{\frac{1}{2}}, \tag{3.18}$$

or

$$\sum_{i=1}^{N} g(A_i) = \sum_{i=1}^{N} (A_i)^{\frac{1}{2}} - K^{\frac{1}{2}}, \tag{3.19}$$

where $K = \sum_{i=1}^{N} A_i$, and $g(\cdot)$ is a function of A_i

For a particular sum, or total throughput K, the function

$$g(A_i) = (A_i)^{\frac{1}{2}} - K^{\frac{1}{2}}$$

is a concave function (since its second derivative is everywhere negative--for positive A_i). It is also a monotonicly increasing function. Nemhauser [15, pp. 53-55] shows that for an arbitrary monotonicly increasing <u>convex</u> function, $\bar{g}(A_i)$, the solution of the problem, minimum $\{\sum_i \bar{g}(A_i) \mid \sum_i A_i \geq K, A_i \geq 0\}$, is achieved at $A_i = K/A_i$, for all i. Hence, for the <u>concave</u> function, $g(A_i)$, the <u>maximum</u> is achieved by letting $A_i = K/A_i$, for all i. In other words, the maximum deviation of approximate from actual inventory cost occurs when the warehouse throughput of a product line is equally divided among the number of items making up that line.

When this situation exists, the following relations are true,

$$\sum_{i=1}^{N} (A_i)^{\frac{1}{2}} = N \cdot (A)^{\frac{1}{2}}, \tag{3.20}$$

and

$$\left(\sum_{i=1}^{N} A_i\right)^{\frac{1}{2}} = (N \cdot A)^{\frac{1}{2}},$$

where $A = \sum_{i=1}^{N} A_i / N$,

N = number of terms in the sum series.

The appropriate multiplier, N_m, to use in this situation to force the approximate costs to equal the actual costs can be derived from the expression,

$$N \cdot (A)^{\frac{1}{2}} = N_m \cdot (N \cdot A)^{\frac{1}{2}}. \tag{3.21}$$

that is,

$$N_m = N \cdot (A)^{\frac{1}{2}} \Big/ (N \cdot A)^{\frac{1}{2}},$$

or,

$$N_m = N^{\frac{1}{2}}.$$

In terms of the distribution model this multiplier will be the square root of the number of items, $\bar{m}$, in product line m. That is, in expression (3.15), N_m = the square root of $(N\bar{M})_m$. The expressions utilizing this formulation of the product multiplier are

a)
$$IC^{P}_{APP} = R^{av}_{j} \cdot \left[\sum_{m=1}^{NM} \left[(N\bar{M})_m^{\frac{1}{2}} \cdot (\tfrac{1}{2}) \cdot C_{jm} \cdot (D_{jm})^{\frac{1}{2}} \right] \right] ,$$

and

b)
$$IC^{L}_{APP} = R^{L}_{j} \cdot \left[\sum_{m=1}^{NM} (N\bar{M})_m^{\frac{1}{2}} \cdot C_{jm} \cdot (D_{jm})^{\frac{1}{2}} \right] ,$$

where

$$D_{jm} = \sum_{\bar{m}=1}^{(N\bar{M})_m} D_{jm\bar{m}} .$$

Note that this formulation implies that the maximum deviation of approximate from true costs actually occurs.

To summarize, the actual warehouse throughputs of individual items in the distribution system result in inventory costs which deviate from that represented in the model. These deviations will be some amount between the minimums and maximums found above. Hence, multipliers, N_m, in the range of 1 to $(N\bar{M})_m^{\frac{1}{2}}$ should be used in the model to correct for the understatement of costs. However, sensitivity analysis should be performed after the initial solution and the value of the multipliers updated.

Constraint Set

There are three restrictions that the optimum distribution system must meet, and which must therefore be incorporated into the model. First, all customer demand for all items must be met. That is,

$$\sum_{i}\sum_{w} X_{ikmw} + \sum_{j}\sum_{w} X_{jkmw} = d_{km}, \text{ for all } k, m. \tag{3.22}$$

Since individual customers are grouped into customer zones in the model, this first constraint implies that the annual demand for each product associated with each customer zone must be delivered by some supply path. This restriction holds for all individual products or items. However, since the model is based on the use of product lines this constraint implies that the composite annual demand of all items within each product line be satisfied. However, only "feasible" supply paths may be used to make the required delivery of a customer zone's demand.

Feasible supply paths are those paths which have an associated lead time equal to or less than an imposed service or delivery time limit. This service time requirement forms the second model constraint. After the supply source (a plant or warehouse) receives a customer's order, that order must be delivered within a specified time limit. This means that the lead time of administration and transportation services associated with the supply path used must be less than or equal to the service time limit. Any supply path meeting this constraint for a particular customer (that is, customer zone) is called a feasible supply path for that customer.

This second restriction is incorporated into the model as an external constraint on the supply path input data. Specifically, infeasible paths are screened out and assigned a large penalty cost as their transportation costs. Feasible paths are assigned their actual transportation costs. Handling the service limit constraint in this manner allows alteration of the time limit in successive model solutions so that the desirability (change in total system's costs) of such alterations can be noted.

The third system constraint is that the production capacity of each plant in the system cannot be exceeded. That is,

$$\sum_k \sum_w X_{ikmw} + \sum_j \sum_w X_{jkmw} \le CAP_{im}, \text{ for all } i, m. \tag{3.23}$$

The model requires that these capacity levels be expressed in the same units as those of customer demand. If demand is expressed in pounds, then plant capacity should be expressed in pounds. This requirement may necessitate use of conversion factors based on a priori, assumed production allocations. That is, production capacity may only be available in units such as total man-hours per year. Customer demand may be expressed in pounds per year. In order to make the two quantities compatible, total manhours must be converted into total pounds (of product).

Normally, the conversion factor of pounds per man-hour is dependent on the product(s) being produced at and shipped from a plant, that is, on the product mix of production. Therefore, in order to obtain the required conversion factor it may be necessary to assume a value of the product mix at each plant in the system and then determine production capacities.

After the model has been solved, the actual product mix at each plant can be used to recompute capacities and the model can be resolved. This process should lead to the appropriate production capacity levels. This same type of analysis may be applied to the determination of unit production cost, if necessary.

It should be pointed out that there are no capacity requirements placed on warehouses. Normally in cities where distribution warehouses are located, warehousing space is, for all practical purposes, unlimited. This applies to both public and leased warehousing. The obvious increase in unit costs of space as the demand for such space increase toward the limits of its supply is assumed to be insignificant.

Basic Assumptions. There are several assumptions underlying the model which may or may not limit its applicability to a specific distribution system. Some of these have been mentioned previously. The purpose of these assumptions is to maintain a realistic representation of the physical distribution system while creating a model which can actually be solved. The most complicated, representative model possible would be useless if it could not be solved.

In order to limit the scope of the analysis, the assumption is made that the number, locations, capacities, and production functions of all manufacturing facilities in the system are fixed. This eliminates the need for explicit consideration in the model of factors such as raw material sources and production schedules.

A second assumption it that the inventory policy used at all warehouses in the system is based on a fixed-quantity reorder strategy. Specifically, this policy is assumed to be the so-called simple "EOQ" policy, with no possibility of shortages. A formal inventory policy for warehouse operations is needed to formulate model components such as average and maximum inventory levels. Since the simple EOQ policy is a common, basic strategy and is not sensitive to deviations from optimal quantities, its use if justifiable.

Another assumption utilized in the model is that warehouse capacity is unlimited. As stated earlier, this holds for both leased and public warehousing. The validity of this assumption is usually upheld in practice, but its use is not critical to the model's formulation or solution.

An inherent assumption in the use of customer zones is that if service of a specific transportation mode is available to a zone (that is, to the zone's key city), it is available to all individual customers within the zone. Further, there is no additional cost associated with obtaining this service for any individual customer. In reality, this assumption is probably violated. The extent and significance of this violation is dependent on the practical application at hand.

A fifth assumption is that all individual customer demand is mutually independent. This allows a forecast to be made of a customer's future demand based on his past demand history, without regard to the future demand of his neighbor. The impact of this assumption on the firm's forecasting subsystem, as well as other forecast system characteristics, is deemed beyond the scope of this research.

The assumption is also being made that the customer demand data given as input to the problem is deterministic in nature.

As stated earlier, the model also assumes that resupply of warehouses can, under normal operating conditions, take advantage of the least-cost tariff rate for any transportation mode. This assumption allows transportation costs along resupply paths to be expressed as a linear function of the volume shipped. The associated marginal rate is the least-cost rate in the tariff schedule of the mode under consideration. Due to the facts that (a) reorder quantities may ordinarily be of sufficient size to warrant the least-cost tariff rate, or, if not, that (b) sufficient resupply quantities can be "built-up" without affecting system operations, this assumption is probably valid in most large-scale distribution systems.

The formulation of the model is also based on the assumption that the safety stock and reorder point levels at a warehouse can be expressed as a linear function of the annual demand satisfied by that warehouse. The functions relating SS_{jm} (safety stock of product m) and RP_{jm} (reorder point of product m) to D_{jm} (annual throughput of warehouse j of product m) are*

$$SS_{jm} = A \cdot (D_{jm}) , \tag{3.24}$$

and

$$RP_{jm} = B \cdot (D_{jm}) ,$$

where A, B = proportionality constants.

Since SS_{jm} and RP_{jm} are dependent on resupply lead time there will be a separate set of constraints (A, B) for each resupply path. This fact forces the SS_{jm} and RP_{jm} functions to be expressed as

---

*Note that these functions imply a zero intercept.

$$SS_{jm} = SS^{w}_{ijm} \cdot (D_{jm}) , \qquad (3.25)$$

and

$$RP_{jm} = RP^{w}_{ijm} \cdot (D_{jm}),$$

where SS^{w}_{ijm} = the proportionality constant between SS_{jm} and D_{jm} when the lead time is defined by the resupply path of plant i to warehouse j by mode w.

RP^{w}_{ijm} = similar definition.

The validity of this assumption that both SS_{jm} and RP_{jm} are linear functions of the annual throughput of a warehouse is dependent on the physical distribution system at hand. Examples can be cited which uphold the assumption, and others which invalidate it. The exact underlying models of SS_{jm} and RP_{jm} are not critical in the formulation and solution of the distribution model. A linear model for each function has been chosen for convenience. If examination of the data in a particular application warrants use of a different underlying model for the two functions, the distribution model can easily be reformulated. For example, nonlinear terms can be handled in a manner similar to the other nonlinear throughput terms, Q_{jm}, resulting for average and maximum inventory expressions.

A final underlying assumption involves unit production costs. Recall that storage costs at public warehouses and leasing fees at leased warehouses are based on the inventory reorder quantity, Q_{jm}. This quantity, in turn, is based on resupply ordering cost, C^{o}_{j}, and unit holding cost C_{m}. Holding cost is some percentage multiplied by the unit production cost. Obviously, the production cost is associated with the resupply plant at which the product m was manufactured.

The assumption being made is that this resupply plant is the plant which is the source of the least-cost resupply path for the warehouse of interest (path costs are based on transportation plus production costs). This assumption will obviously hold except in limited cases where the capacity of a plant is exceeded. Then, a plant other than the least-cost plant must be utilized in the warehouse's resupply path.

Specific Model

The general model discussed previously can be translated into specific, quantitative relations and expressed as follows:

Let

$i \Rightarrow$ index referring to plants,

$j \Rightarrow$ index referring to warehouses,

$p \Rightarrow$ superscript referring to public type warehouses,

$L \Rightarrow$ superscript referring to leased type warehouses,

$J_p \Rightarrow$ the subset of public type warehouses,

$J_L \Rightarrow$ the subset of leased type warehouses,

$k \Rightarrow$ index referring to customers,

$m \Rightarrow$ index referring to products,

$w \Rightarrow$ index referring to transportation modes,

X_{ijmw} = the flow of product m from plant i to warehouse j using mode w,

X_{ikmw} = the flow of product m from plant i to customer k via mode w,

X_{jkmw} = The flow of product m from warehouse j to customer k via mode w,

S_{ijmw} = the unit shipment cost (transportation + production) of product m along resupply path i-j-w,

r_{ikmw} = the unit shipment cost (transportation + production) of product m along supply path i-k-w,

C_{jkmw} = the unit transportation cost of product m along supply path j-k-w,

R_j^{in} = the unit handling charge at public warehouse j,

R_j^{av} = the unit storage charge at public warehouse j,

R_j^{L} = the unit leasing fee at leased warehouse j,

FC_j = the fixed cost component of leased warehouse j's operating cost,

C_j^{o} = the cost of placing a resupply order at warehouse j,

C_{i*j}^{m} = the unit inventory holding cost (a percentage multiplied by unit production cost of product m at j's least-cost resupply plant i),

SS_{ijm}^{w} = the proportionality constant for the safety stock of product m at warehouse j when the resupply path is i-j-w,

RP_{ijw}^{w} = the proportionality constant for the reorder point level of product m at warehouse j when the resupply path is i-j-w,

N_m = the product multiplier for product m,

d_{km} = the annual demand of customer zone k for product m,

CAP_{im} = the annual capacity of plant i for manufacturing product m,

Y_j = a zero, one decision variable.

minimize

$$\sum_i \sum_j \sum_m \sum_w S_{ijmw} X_{ijmw} \quad \left\{ \begin{array}{l} \text{Total plant to warehouse} \\ \text{transportation costs} \end{array} \right. \tag{3.26}$$

$$+ \sum_i \sum_k \sum_m \sum_w r_{ikmw} X_{ikmw} \quad \left\{ \begin{array}{l} \text{Total plant to customer} \\ \text{"direct" transportation costs} \end{array} \right.$$

$$+ \sum_j \sum_k \sum_m \sum_w C_{jkmw} X_{jkmw} \quad \left\{ \begin{array}{l} \text{Total warehouse to customer} \\ \text{transportation costs} \end{array} \right.$$

$$+ \sum_{j \in J_p} R_j^{in} \cdot \left(\sum_k \sum_m \sum_w X_{jkmw} \right) \quad \left\{ \begin{array}{l} \text{Total handling costs at} \\ \text{public type warehouses} \end{array} \right.$$

$$+ \sum_{j \in J_p} \left[12 \cdot R_j^{in} \cdot \left[\sum_m N_m \cdot \left(\tfrac{1}{2} \left(\frac{2C_j^o}{C_{i*j}^m} \right)^{\frac{1}{2}} \right) \left(\sum_k \sum_w X_{jkmw} \right)^{\frac{1}{2}} \right] \right]$$

$$+ \sum_i \sum_m \sum_w \sum_{j \in J_p} (12 \cdot R_j^{in}) (SS_{ijm}^w) X_{ijmw} \quad \left\{ \begin{array}{l} \text{Total storage costs} \\ \text{at public type warehouses} \end{array} \right.$$

$$+ \sum_{j \in J_L} FC_j Y_j \quad \left\{ \begin{array}{l} \text{Total of the fixed cost components of} \\ \text{operating costs at leased type warehouses} \end{array} \right.$$

$$+ \sum_{j \in J_L} R_j^L \cdot \left[\sum_m N_m \cdot \left(\frac{2C_j^o}{C_{i*j}^m} \right)^{\frac{1}{2}} \left(\sum_k \sum_w X_{jkmw} \right)^{\frac{1}{2}} \right]$$

$$+ \sum_i \sum_m \sum_w \sum_{j \in J_L} (R_j^L)(RP_{ijm}^w) X_{ijmw} \quad \left\{ \begin{array}{l} \text{Total leasing fees at} \\ \text{leased type warehouses} \end{array} \right.$$

subject to:

* (a) $\sum_{i}\sum_{w} X_{ikmw} + \sum_{j}\sum_{w} X_{jkmw} = d_{km}$,

for all k, m

Total flow of product m along all supply paths having k as a sink must equal k's demand for product m

(b) $\sum_{k}\sum_{w} X_{ikmw} + \sum_{j}\sum_{w} X_{ijmw} \leq CAP_{im}$,

for all i, m

Total flow of product m out of plant i must be less than or equal to the production capacity of i for m

(c) $\sum_{i}\sum_{w} X_{ijmw} = \sum_{k}\sum_{w} X_{jkmw}$,

for all j, m

Total flow of product m into warehouse j must equal total flow out

(d) $X_{ijmw}, X_{ikmw}, X_{jkmw} \geq 0$,

for all i, j, k, m, w

Auxiliary constraints

(e) $$Y_j = \begin{cases} 0, & \text{if } \sum_{k}\sum_{m}\sum_{w} X_{jkmw} = 0, \\ 1, & \text{if } \sum_{k}\sum_{m}\sum_{w} X_{jkmw} \geq 0, \end{cases}$$

for all $j \in J_L$.

Auxiliary constraints

---

* The set of i, w and j, w for which the terms in constraint (a) are summed over are those corresponding to feasible supply paths--those meeting the service limit. All other constraints are summed over all possible values of the indices.

CHAPTER IV

SOLUTION PROCEDURE

Basic Solution Approach

The model (3.26) developed in the preceding chapter constitutes the programming problem of minimizing a strictly concave, mixed integer, zero-one function over a convex feasible region. There is no method currently available which will yield an exact, globally optimal solution to this problem. However, a locally optimal solution can be obtained by performing several preliminary analyses to reduce the complexity of the model, and then executing the following two-phase solution procedure.

Phase One

(1) Determine an effective starting point of warehouse throughputs.

(2) Successively iterate through a "linearization-solution" routine until either the value of the objective function is greater than its predecessor, or two successive solutions repeat themselves. The routine is as follows:

a) using the current set of warehouse throughputs, transform the nonlinear objective function into a linearly equivalent function;

b) solve this linearized model;

c) use the solution to calculate a new set of warehouse throughputs;

d) return to step a).

(3) Using additional starting points, repeat the linearization-solution routine.

(4) Let the least-cost solution resulting from use of the various starting points be the end result of Phase One.

Phase Two

(1) Formulate unit penalty costs based on the results of Phase One.

(2) Attach these penalties to those warehouse-related supply paths which would seem to be unprofitable in the optimal problem solution.

(3) Using the starting point of the penalized Phase One results, iterate the linearization-solution routine until the stopping criteria is met.

(4) The best of the Phase One and the Phase Two results is the final solution to the problem.

In other terms, the successive linearization procedure used to solve the distribution model consists of four steps. First, external analyses are performed to estimate certain parameters of the model, and to reduce the model's complexity. Second, the non-linear model is transformed into a linear programming model by factoring out the nonlinear terms in the objective function and using approximate values for these terms to obtain a linear approximation. Since the constraints set is originally linear, no transformation need be performed on it.

Third, an initial value for each of the nonlinear terms in the model is determined (these terms are the square roots of warehouse throughputs). This set of initial values is based on warehouse throughputs associated with an a priori "good" model solution. As the first of two parts of this third step, the linearized model based on the initial throughput set is solved and a new throughput set is determined from the initial solution. The original, nonlinear model is again linearized--using this second throughput set. This linearization-solution process is repeated until two successive solutions are the same. This is the end of part one. Part two entails repeating the linearization-solution process based on a series of additional starting points.

The fourth solution step, called "Phase Two," starts with a throughput set, or starting point, obtained from the best solution found in the previous step (Phase One). A set of unit penalty costs is formulated and attached to those supply paths in this starting point solution which are deemed to be uneconomical. The purpose of these penalty costs is to adjust for a specific bias in the first phase results. After the penalty costs are attached, the linearization-solution process is undertaken again. The best of the Phase One and Phase Two results is the solution to the problem.

Before explaining each step in detail, it may be helpful to define (or redefine) the following terms used in the description of the solution process.

(1) A source is a plant or warehouse forming the initial part of a supply or resupply path.

(2) A sink is the end destination (customer-product, or warehouse) of either a supply path or a resupply path.

(3) A customer-sink is a customer-product combination which serves as the demand end of a supply path. Note that if NK is the number of customer zones in the model, and NM is the number of product lines, there will be NK·NM sinks in the model associated with supply paths.

(4) A feasible supply path is a supply path which has an associated transit plus administrative lead time equal to or less than the required service time limit.

(5) A linearized function is a nonlinear function which has been transformed into a linear function based on assumed values of certain decision variables, such as warehouse throughputs.

(6) A solution is a set of values of the decision variables in the model. These variables are the flows of each product over all supply and resupply paths in the system.

(7) A throughput set is the set of annual volumes or tonnage of each product which flows through each warehouse. A throughput set is derived from the solution to the model. In other words, a throughput set is the set of total customer demand for each product assigned to each warehouse.

(8) A starting point is a particular throughput set.

External Analyses

The first step in the solution procedure is to reduce the original model to a less complex form. This can be done by (a) handling the service-limit constraint as an external system constraint, (b) eliminating from the model explicit consideration of the transportation mode alternatives and (c) formulating safety stock and reorder point levels as a function of annual warehouse throughput.

Service Limit

As mentioned previously and as indicated in the general model of Chapter III (3.26), the service or supply time limit can be used externally to screen out those supply paths for each customer which are not feasible paths. A "large" unit cost is assigned to each infeasible path as its transportation cost. In this manner, all supply paths explicitly appear in the model--feasible paths having normal unit costs, infeasible ones having "large" unit costs. The favorable economics of utilizing a feasible path as opposed to an infeasible one as a customer's supply path will prevent infeasible paths from being part of the model's solution.

Supply Mode

Model (3.26) can be reduced to a less complicated form by eliminating the need for the mode subscript (w) which appears in connection with supply paths and with resupply paths. Along supply paths, transportation costs for alternative modes are based on the average order size of the customer sink associated with each path. Specifically, each transportation mode between any source and sink (that is, along any path) has a tariff rate schedule in which rates are a function of shipment sizes. The exact rate applicable to a shipment along a supply path is determined by the size of that shipment. Since a customer's average order size (for a particular product) represents the expected value of all sizes of shipments to that customer, the average order size can be used to determine the average unit transportation rate of all shipments to that customer from any source. That is, given the tariff rate schedule associated with any mode for shipments between any source and sink, and the expected shipment size to that sink, a unit cost can be determined which represents the average unit transportation cost of all shipments along that path.

Therefore, once the average shipment size of each product to all customers has been estimated, a matrix of unit supply path transportation costs can be determined. This matrix would represent the unit cost, t^{w}_{ikm} and t^{w}_{jkm}, associated with each alternate transportation mode for all combinations of supply sources and sinks. Based on this matrix, the mode, w^{*}_{ikm}, to use for deliveries of product m from sink i to source k is

$$w^{*}_{ikm} = \left\{ w \;\middle|\; t^{w*}_{ikm} = \underset{w}{\text{minimum}} \, (t^{w}_{ikm}) \right\} . \tag{4.1}$$

Similarly, for deliveries of m from j to k use w^{*}_{jkm} such that

$$\left\{ w^*_{jkm} = w \mid t^{w*}_{jkm} = \underset{w}{\text{minimum}} \, (t^w_{jkm}) \right\} . \tag{4.2}$$

Resupply Mode

As in supply paths, tariff rates for all resupply paths are dependent on the size of shipment transported along the path. The shipment quantities transported along resupply paths are the warehouse reorder quantities, Q_{jm}. However, as stated previously, the assumption is made that the resupply subsystem is such that the least-cost tariff rate for each transportation mode can be utilized. Unfortunately, the mode offering the smallest rate between a supply source and sink cannot be selected without further investigation.

As mentioned earlier, there is a trade-off of costs associated with the use of alternative transportation modes and the costs of carrying inventory. To reiterate, the more economical transportation modes are normally slower, forcing a longer resupply lead time for a specific warehouse. In order to protect against shortages during this addition lead time, the warehouse must carry a larger buffer or safety stock. Since the average and maximum inventory levels increase with larger safety stocks, the cost of carrying inventory at the warehouse increases.

Therefore, the mode to use between any resupply source (plant) and sink (warehouse) is that mode which minimizes the combined transportation and warehouse operating costs. As indicated previously, the appropriate transportation costs to compare are the least-cost tariff rates, t^w_{ijm}. In order to determine the appropriate warehouse operating costs to compare, examine the total annual cost expression, TC^w_{ijm}, for <u>each</u> resupply path, i-j-w, for terms influenced by use of alternative modes. This expression for <u>public</u> warehousing is*

$$TC^w_{ijm} = \left[12 \cdot R^{av}_j \cdot \left[(N_m \cdot C_{jm} \cdot (\sum_i \sum_w X_{ijmw})^{\frac{1}{2}} + SS^w_{ijm} X_{ijmw})\right]\right. \tag{4.3}$$

$$\left. + R^{in}_j X_{ijmw} \right] \qquad \Big\{ \text{Warehouse operating costs}$$

$$+ \left[t^w_{ijm} X_{ijmw} \right] . \qquad \Big\{ \text{Transportation cost}$$

Disregarding all terms not affected by changes in w, the total cost expression can be reduced to

$$TC^w_{ijm} = 12 \cdot R^{av} \cdot (SS^w_{ijm} X_{ijmw}) + t^w_{ijm} X_{ijmw}. \tag{4.4}$$

Factoring out the constant X term, the expression (4.4) can be rewritten as

$$TC^w_{ijm} = 12 \cdot R^{av}_j \cdot (SS^w_{ijm}) + t^w_{ijm} . \tag{4.5}$$

The transportation mode to use for resupply of product m from source i to sink j is that mode, w^*_{ijm}, which minimizes the above expression. That is, for $j \epsilon J_p$,

---

*Note than an equivalent method of determining warehouse throughput is being used; that is, demand is summed over i rather than k. Note also that $t^w_{ijm} = t^w_{ij}$ for all m.

$$w^{*}_{ijm} = \left\{ w \;\middle|\; TC^{w*}_{ijm} = \underset{w}{\text{minimum}} \, (12 \cdot R^{av}_{j} \cdot SS^{w}_{ijm} + t^{w}_{ijm}) \right\} . \quad (4.6)$$

Similarly, for $j \in J_L$, the transportation mode w^{*}_{ijm} to use to resupply the inventory level of product m of warehouse j when the resupply plant is i, is

$$w^{*}_{ijm} = \left\{ w \;\middle|\; TC^{w*}_{ijm} = \underset{w}{\text{minimum}} \, (R^{L}_{j} \cdot RP^{w}_{ijm} + t^{w}_{ijm}) \right\} . \quad (4.7)$$

Safety Stock and Reorder Point Coefficients

An assumption underlying the model is that the amount of safety stock (SS_{jm}) and the reorder point level (RP_{jm}) are both linear functions of annual warehouse throughput. Further, these functions are assumed to be represented by the zero-intercept models

$$SS_{jm} = SS^{w}_{ijm} \cdot D_{jm} , \quad (4.8)$$

and

$$RP_{jm} = RP^{w}_{ijm} \cdot D_{jm} .$$

The constants of proportionality in these models must be estimated before either a) resupply transportation modes can be determined, or b) the distribution model (3.26) can be solved. One approach to obtaining these estimates is to generate data points of safety stock and reorder point levels by applying the definition of these terms to the annual demand data used in the model. After generating a sufficient number of data points, least-square estimators of the constants or proportionality can be obtained through regression analysis.

In other words, by determining the amount of safety stock and the reorder point level associated with a specific lead time and various annual demands (or warehouse throughputs), a distribution of safety stock and reorder point values vs. annual demand values can be generated. Then, the line of "best-fit" can be determined relating safety stock (and reorder point) values to annual demands. The slope parameter of this line is the required safety stock (and reorder point) coefficient, SS^{W}_{ijm} (and RP^{W}_{ijm}).

The steps in this process of determining the required coefficients are outlined in the following list.

(1) Determine all possible values of resupply lead times (in days).

(2) Form a distribution of daily demand for the first customer zone's demand for the first product (the total demand associated with this distribution represents the first possible value of annual warehouse throughput, D_1).

(3) For each value of lead time (LT),

a) search the distribution of daily demand to find the maximum demand during LT, that is, the reorder point level, RP^{LT}_{D}. Specifically,*

$$RP^{LT}_{D} = \text{maximum (DDLT)} ,$$

where DDLT = the demand during lead time,

RP^{LT}_{D} = the reorder point level required to meet an annual demand D when the resupply lead time is LT.

---

*Note that reorder point is normally defined as the maximum "reasonable" demand lead time, and "reasonable" is here assumed to imply a "0" probability of stockout.

b) Search the distribution to find the average demand during lead time. The difference between this average and RP_D^{LT} is the required amount of safety stock to carry for the annual demand associated with this distribution. That is,

$$SS_D^{LT} = RP_D^{LT} - \overline{DDLT},$$

where $\overline{DDLT}$ = average DDLT,

SS_D^{LT} = amount of safety stock required to meet an annual demand D when the resupply lead time is LT.

Assuming that the expressions for SS_D^{LT} and RP_D^{LT} are based on only one year's (365 days) demand data, the expression for $\overline{DDLT}$ is

$$\overline{DDLT} = \frac{\sum_{i=1}^{LT-1} i \cdot (d_i) + \sum_{i=LT}^{365-LT+1} LT \cdot (d_i) + \sum_{i=365-LT+2}^{365} (365-i+1)\ (d_i)}{(365-LT+1)}$$

where d_i = demand for day i.

For example, if the value of the lead time is three days (LT=3) then

$$\overline{DDLT} = \left[(d_1+d_2+d_3)+(d_2+d_3+d_4)+(d_3+d_4+d_5)+---+(d_{362}+d_{363}+d_{364}) + (d_{363}+d_{364}+d_{365}) \right] \Big/ (363)$$

(4) Form a second daily demand distribution based on the first customer's demand for the second product (the total of this new distribution represents a second possible throughput value, D_2).

(5) Search this new distribution for the values of RP and SS for each value of LT. That is, repeat step (3).

(6) For all possible combinations of customer-sink demand (that is, all possible throughputs, D_i, a warehouse might experience), form a new distribution, search it, and determine SS and RP. That is, repeat steps (4) and (5).

For each value of lead time which might be associated with a resupply path, these first six steps will result in the required amount of safety stock and the reorder level to employ for all possible values of warehouse throughputs. Assuming that the zone demands are deterministic in nature (an implicit model assumption), steps one through six have completely enumerated all possible values of safety stock and reorder point.* Hence, by regressing safety stock (and reorder point) values on annual demand values appropriate proportionality constants can be obtained. These constants allow a value of the required amount of safety stock (and reorder point) to be calculated for any given demand, or warehouse throughput, based on the resupply lead time involved.

The coefficients obtained from this regression analysis are the following least-squares estimators (see [16] for formulation):

a) for LT associated with resupply path i-j-w,

$$\hat{b}_1 = SS^{w}_{ijm} = \frac{\sum_{\ell \epsilon S^m} (SS^{LT}_{D_\ell})(D_\ell)}{\sum_{\ell \epsilon S^m} (D_\ell)^2} \qquad (4.9)$$

---

*It should be pointed out that by basing the safety stock and reorder point data points on <u>each</u> <u>component</u> of the <u>cumulative</u> sum of sink demands--rather than on all possible combination of sink demands--computational time involved would be effectively reduced without the loss of a significant amount of accuracy.

where S^m = the set of all ℓ for which D_ℓ involves sink demand for product m,

and,

b) similarly,

$$\hat{b}_2 = RP^w_{ijm} = \frac{\sum_{\ell \epsilon S^m} (RP^{LT}_{D_\ell})(D_\ell)}{\sum_{\ell \epsilon S^m} (D_\ell)^2} \quad . \tag{4.10}$$

To reiterate, the safety stock and reorder point coefficients in model (3.26)-- SS^w_{ijm} and RP^w_{ijm} --can be found by first generating data points based on searching a number of daily demand distributions for the appropriate values of safety stock and reorder point levels. Then, based on these data points, least-squares estimators of the coefficients can be determined.

Model Transformation

Utilizing the external analyses described in the previous section, the original model (3.26) can be reduced to the expression

$$\text{minimize } \sum_i \sum_j \sum_m S_{ijm} X_{ijm} + \sum_i \sum_k \sum_m r_{ikm} X_{ikm} + \sum_j \sum_k \sum_m C_{jkm} X_{jkm} \quad \left\{ \begin{array}{l} \text{Total} \\ \text{Trans.} \\ \text{Cost} \end{array} \right. \tag{4.11}$$

$$\left. \begin{array}{l} + \sum_k \sum_m \sum_{j \epsilon J_p} R^{in}_j X_{jkm} + \sum_{j \epsilon J_p} \left[12(R^{av}_j) \sum_m \frac{1}{2}(N_m)(H_{jm})(\sum_k X_{jkm})^{\frac{1}{2}} \right] \\ \qquad + \sum_i \sum_m \sum_{j \epsilon J_p} \left[12(R^{av}_j)(SS^{w*}_{ijm}) X_{ijm} \right] \end{array} \right\} \begin{array}{l} \text{Public} \\ \text{Warehouse} \\ \text{Operating} \\ \text{Costs} \end{array}$$

$$+\sum_{j\epsilon J_L} FC_j Y_j + \sum_{j\epsilon J_L} \left[(R_j^L) \sum_m (N_m)(H_{jm}(\sum_k X_{jkm})^{\frac{1}{2}} \right]$$

$$+\sum_i \sum_m \sum_{j\epsilon J_L} \left[(R_j^L)(RP_{ijm}^{w*}) X_{ijm} \right]$$

Leased Warehouse Operating Costs

subject to

a) $$\sum_i X_{ikm} + \sum_j X_{jkm} = d_{km} ,$$

for all k, m,

Customer Demand Constraints

b) $$\sum_k X_{ikm} + \sum_j X_{ijm} \le CAP_{im},$$

for all i, m,

Plant Capacity Constraints

c) $$\sum_i X_{ijm} = \sum_k X_{ikm} ,$$

for all j, m,

Warehouse Throughput Constraints

d) $$X_{ijm}, X_{ikm}, X_{jkm} \ge 0,$$

for all i, j, k, m,

Nonnegativity Constraints

e) $$Y_j = 0, \text{ or } 1,$$

for all $j\epsilon J_L$,

Zero-One Variable Constraints

where $$H_{jm} = \frac{2.C_j^o}{C_{i*j}^m}^{\frac{1}{2}}$$

This nonlinear model can be transformed into an equivalent form which is linear in "X" when a value for the throughput set is assumed. Essentially, this transformation involves factoring out, or dividing through the terms nonlinear in "X". The transformation proceeds as follows:

(a) substitute the following terms into the model:

$$S^P_{ijm} = \left[S_{ijm} + 12(R^{av}_j)(SS^{w*}_{ijm})\right], \text{ for all } j\epsilon J_p,$$

$$S^L_{ijm} = \left[S_{ijm} + (R^L_j)(RP^{w*}_{ijm})\right], \text{ for all } j\epsilon J_L,$$

$$A^P_{jkm} = (C_{jkm} + R^{in}_j), \text{ for all } j\epsilon J_p,$$

$$A^L_{jkm} = (C_{jkm}), \text{ for all } j\epsilon J_L,$$

$$B^P_{jm} = 12\cdot(R^{av}_j)(N_m)(\tfrac{1}{2}H_{jm}), \text{ for all } j\epsilon J_p,$$

$$B^L_{jm} = (R^L_j)(N_m)(H_{jm}), \text{ for all } j\epsilon J_L,$$

(b) Rewrite the objective function using these substitutions,

$$\text{minimize} \sum_i \sum_k \sum_m r_{ikm} X_{ikm} + \sum_i \sum_m \sum_{j\epsilon J_p} S^p_{ijm} X_{ijm} \tag{4.12}$$

$$+ \sum_i \sum_m \sum_{j\epsilon J_L} S^L_{ijm} X_{ijm}$$

$$+ \sum_k \sum_m \sum_{j\epsilon J_p} A^P_{jkm} X_{jkm} + \sum_k \sum_m \sum_{j\epsilon J_L} A^L_{jkm} X_{jkm}$$

$$+ \sum_m \sum_{j\epsilon J_P} B^P_{jm} \cdot \left(\sum_k X_{jkm}\right)^{\frac{1}{2}} + \sum_m \sum_{j\epsilon J_L} B^L_{jm} \cdot \left(\sum_k X_{jkm}\right)^{\frac{1}{2}}$$

$$+\sum_{j\epsilon J_L} FC_j Y_j \, .$$

(c) Using the fact that for a given m, j, $\sum_k X_{jkm} = \sum_{k*} X_{jk*m}$, where k = 1, 2, -----N and k* = 1, 2, -----, N, and that likewise, $\sum_k \sum_m X_{jkm} = \sum_{k*} \sum_{m*} X_{jk*m*}$, for a given j, the following unity expressions can be formulated:

$$\frac{\sum_k X_{jkm}}{\sum_{k*} X_{jk*m}} = 1, \text{ for a given j, m,}$$

and

$$\frac{\sum_k \sum_m X_{jkm}}{\sum_{k*} \sum_{m*} X_{jk*m*}} = 1, \text{ for a given j.}$$

Letting $D_{jm} = \sum_{k*} X_{jk*m}$, and $D_j = \sum_{k*} \sum_{m*} X_{jk*m*}$, and specifying D_{jm} (or D_j) = one if $\sum_{k*} X_{jk*m}$ = (or $\sum_{k*} \sum_{m*} X_{jk*m*}$) = zero, these expressions can be written as

$$\frac{\sum_k X_{jkm}}{D_{jm}} = 1, \qquad \text{and} \qquad \frac{\sum_k \sum_m X_{jkm}}{D_j} = 1. \tag{4.13}$$

(d) Multiply the nonlinear terms in the objective function (4.12) by these expressions (4.13). The result is

$$\text{minimize} \sum_i \sum_k \sum_m r_{ikm} X_{ikm} + \sum_i \sum_m \sum_{j \epsilon J_p} S^P_{ijm} X_{ijm} + \sum_i \sum_m \sum_{j \epsilon J_L} S^L_{ijm} X_{ijm} \tag{4.14}$$

$$+ \sum_m \sum_{j \epsilon J_P} A^P_{jkm} X_{jkm} + \sum_k \sum_m \sum_{j \epsilon J_L} A^L_{jkm} X_{jkm}$$

$$+ \sum_m \sum_{j \epsilon J_P} B^P_{jm} \cdot (\sum_k X_{jkm})^{\frac{1}{2}} \left[\frac{\sum_k X_{jkm}}{D_{jm}} \right] + \sum_m \sum_{j \epsilon J_L} B^L_{jm} \cdot (\sum_k X_{jkm})^{\frac{1}{2}} \left[\frac{\sum_k X_{jkm}}{D_{jm}} \right]$$

$$+ \sum_{j \epsilon J_L} (FC_j) \left[\frac{\sum_k \sum_m X_{jkm}}{D_j} \right] .$$

(e) Transpose the terms of (4.14) and express the objective function as

$$\text{minimize} \sum_i \sum_k \sum_m r_{ikm} X_{ikm} + \text{---} + \sum_k \sum_m \sum_{j \epsilon J_L} A^L_{jkm} X_{jkm} \tag{4.15}$$

$$+ \sum_m \sum_{j \epsilon J_P} B^P_{jm} \left[\frac{(\sum_k X_{jkm})^{\frac{1}{2}}}{D_{jm}} \right] \sum_k X_{jkm} + \sum_m \sum_j \sum_{j \epsilon J_L} B^L_{jm} \left[\frac{(\sum_k X_{jkm})^{\frac{1}{2}}}{D_{jm}} \right] \sum_k X_{jkm}$$

$$+ \sum_{j \epsilon J_L} \left[\frac{(FC_j)}{D_j} \right] \sum_k \sum_m X_{jkm}$$

(f) Note the following reduction of terms:

$$\frac{(\sum_k X_{jkm})^{\frac{1}{2}}}{D_{jm}} = \frac{1}{(D_{jm})^{\frac{1}{2}}} \tag{4.16}$$

(g) Using expressions (4.16) re-express (4.15) as

$$\text{minimize} \sum_i \sum_k \sum_m r_{ikm} X_{ikm} + \text{----} + \sum_k \sum_m \sum_{j \epsilon J_L} A^L_{jkm} X_{jkm} \tag{4.17}$$

$$+ \sum_m \sum_{j \epsilon J_P} \left[\frac{B^P_{jm}}{(D_{jm})^{\frac{1}{2}}} \right] \sum_k X_{jkm} + \sum_m \sum_{j \epsilon J_L} \left[\frac{B^L_{jm}}{(D_{jm})^{\frac{1}{2}}} \right] \sum_k X_{jkm}$$

$$+ \sum_{j \epsilon J_L} \left[\frac{FC_j}{D_j} \right] \sum_k \sum_m X_{jkm} \, .$$

(h) Bring constant terms inside the sum signs and combine terms as follows:

$$\text{minimize} \sum_i \sum_k \sum_m r_{ikm} X_{ikm} + \sum_i \sum_m \sum_{j \epsilon J_P} S^P_{ijm} X_{ijm}$$

$$+ \sum_i \sum_m \sum_{j \epsilon J_L} S^L_{ijm} X_{jkm}$$

$$+ \sum_k \sum_m \sum_{j \epsilon J_P} \left[A^P_{jkm} + \frac{B^P_{jm}}{(D_{jm})^{\frac{1}{2}}} \right] X_{jkm}$$

$$+ \sum_k \sum_m \sum_{j \epsilon J_L} \left[A^L_{jkm} + \frac{B^L_{jm}}{(D_{jm})^{\frac{1}{2}}} + \frac{FC_j}{D_j} \right] X_{jkm}$$

Note that the above expression (4.18) is a linear function in "$\underline{X}$" after D_{jm} and D_j values have been determined. This allows (3.26) to be expressed in an equivalent linear form (4.18) when $D_{jm} = \sum_k X_{jkm}$ and $D_j = \sum_k \sum_m X_{jkm}$. Also note that the constraint set of the original model (3.26) is still valid, except that the zero-one constraints on Y_j are no longer required. Therefore the equivalent linearized model is (4.18)

subject to

a) $\sum_i X_{ikm} + \sum_j X_{jkm} = d_{km}$, for all k, m,

b) $\sum_k X_{ikm} + \sum_j X_{ijm} \leq CAP_{im}$, for all i, m

c) $\sum_i X_{ijm} = \sum_k X_{jkm}$, for all j, m,

d) $X_{ijm}, X_{ikm}, X_{jkm} \geq 0$, for all i, j, k, m.

Theoretical Considerations

The two-phase successive linearization solution procedure which utilizes the linearized model (4.18) developed in the previous section is based on recognition of two theoretical considerations.

Local Optimality

The first consideration is that by merely linearizing a nonlinear problem and solving this linear model, either in one pass or successively, there is no guarantee that the final answer will be the globally optimal problem solution. The best that a linearization approach can hope for is to ensure that the final answer is at least a local optimum.

The reason for the possibility of a locally optimal solution was pointed out in the first chapter. Basically, the reason is that there are several extreme points of the convex solution space which are optimal solutions within local areas. The global solution corresponds to one of these local optima. Since the successive linearization solution approach is based on utilizing mathematical programming to find the "best" solution of a linearized objective function, one of this approach's weaknesses is that it has the same "near-sightedness" as mathematical programming. That is, it finds a solution which is locally optimal, but not necessarily the global optimum.

In addition, the procedure may transform a locally optimal extreme point into what appears to be a nonoptimal extreme point. This may occur because the solution procedure transforms the original nonlinear objective function into a linear function. An extreme point may be optimal with respect to the nonlinear function but not to the linear one. Graphically, this phenomena occurs as shown in figure 14.

Note that before the nonlinear function is transformed to a linear form, extreme point A is a local optimum. However, after the original function is linearized, point A is no longer optimal.

The inherent bias of a successive linearization approach mentioned in the second chapter may result from either of these two reasons ("nearsightedness" or transformation). As pointed out by Baumol and Wolfe [10], the solution to a concave location problem by a linearization method tends to utilize more location facilities than is economically desirable. That is, there are too many warehouses in the solution with positive throughputs. An extreme point solution having fewer, larger warehouses and a smaller total cost may exist.

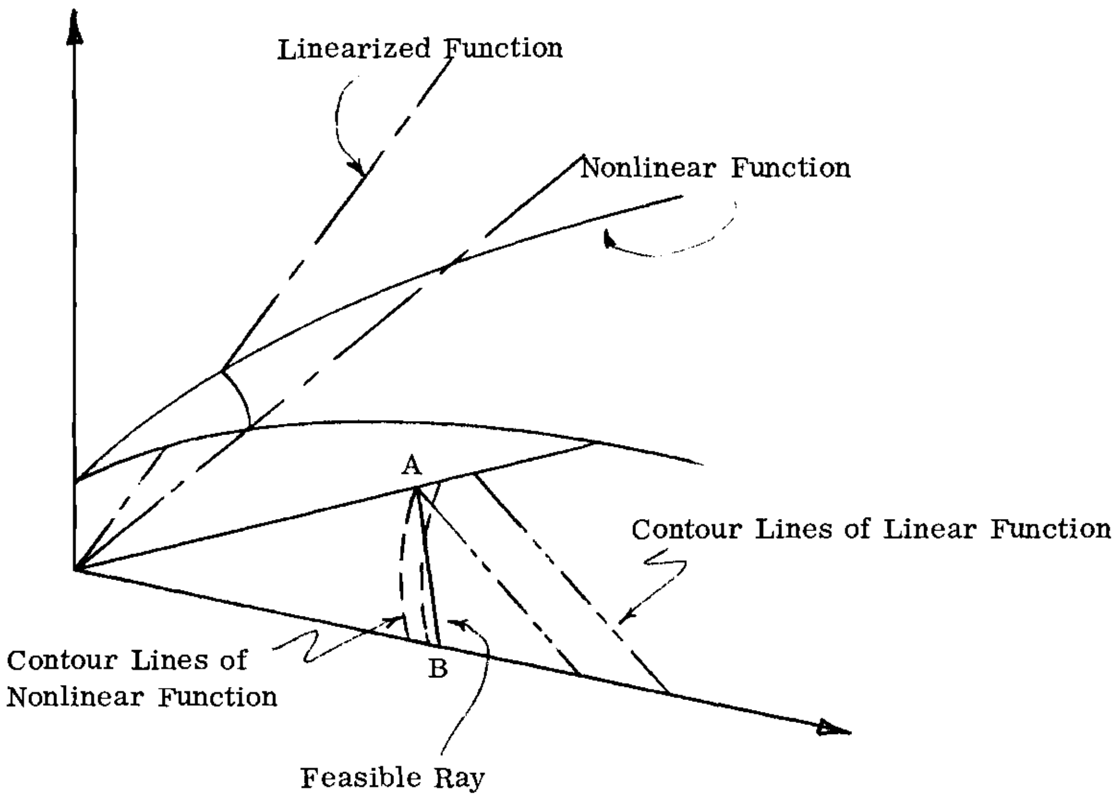


Figure 14. Suboptimality Due to Transformation.

The reason that ordinary successive linearization (such as Baumol and Wolfe's method [10], or Hammond's method [12]) may exhibit this bias and result in a local optimum is that either

a) not enough extreme points are examined in the solution search, or

b) an extreme point associated with use of fewer warehouses which is globally optimal with respect to the nonlinear objective function is transformed into a non-optimal extreme point with regard to the linearized function, and is therefore passed over by the solution search.

Suboptimality

As indicated in Chapter II, a suboptimal solution can result from a successive linearization procedure due to the "crossing" of two or more warehouse operating cost functions. That is, a suboptimal solution can arise if there exist in the model a customer-sink which has two or more supply paths originating at warehouses and the operating cost functions associated with these warehouses intersect (that is, have the same value at some level(s) of warehouse throughput). An example of a location problem with this characteristic and of the successive linearization procedure (Hammond's method [12]) resulting in a suboptimal solution is given in the appendix.

As brought out in the earlier discussion (Chapter II) concerning this particular model feature, the reason that a successive linearization method can run into difficulty in solving this type problem is inherent in the linearization procedure itself. Specifically, the nonlinear objective function is transformed (linearized) using predetermined values for the throughput set. These values may be throughput levels below the point of intersection of two or more operating functions. If the allocation (of sink demand) resulting from a solution based on these throughput values results in new warehouse operating levels above the intersection point, then an incorrect allocation may have been made. In other words, a decision was made involving warehouse operations at operating levels (throughputs) having different cost relationships than the relationships at the levels upon which the decision was based. Such a decision can be suboptimal if the operating cost functions intersect or "cross" between these two operating levels.

An appropriate method for preventing suboptimality in a location problem having this characteristic is to ensure that decisions involving those cost functions which intersect have been based on operating levels on both sides of the point of intersection. A successive linearization procedure may accomplish this without modification. This is because decisions are made based on a starting point and on successive points (levels) in some direction. However, the procedure may terminate before levels beyond an intersection point have been examined.

To guard against this premature termination, additional starting points can be utilized which ensure that all intervals of operating levels of the functions are examined by the successive solution procedure. Providing these additional starting points is one of the objectives of Phase One of the two-phase solution procedure used to solve model (3.26).

Phase One

Linearization-Solution Routine

The basis of the first phase of the solution procedure is an iterative linearization-solution routine. This routine entails transforming the nonlinear model (3.26) into the linear model (4.18) based on a particular starting point (a starting point is a specific throughput set). The specific throughputs used also determine the appropriate level of variable model cost parameters whose values depend on throughput levels. Note that the linear model (4.18) can be expressed in terms of the standard transportation problem.* Therefore, the efficient

---

*This is obvious when the model is thought of in terms of flows along paths from sources to sinks. Specifically, the model can be expressed as

minimize Σ (Unit Cost)(Flow Along All Paths)

subject to Σ (Flow Along All Supply Paths) = (Customer-Sink Demand), (cont'd) solution algorithms available for solving the transportation problem can be used to solve model (4.18). The best media for understanding the equivalence of (4.18) and the transportation problem is the standard transportation array or matrix. Figure 15 shows this matrix. Note that in the matrix there are transshipments allowed among warehouses but nowhere else.

After the linear model has been solved, the next step of the linearization-solution routine is to calculate the resulting set of warehouse throughputs. This set is based on the values of the decision variables (supply and resupply path flows) in the current solution. If this current solution is an improvement over the previous solution, a new linear formulation of the model is derived by calculating new unit cost based on the throughout set just determined. The model is re-solved, and the routine continues.

Given a starting point, the specific steps of this linearization-solution routine can be described in the following outline.

First Stage

(1) Transform the nonlinear model into a linear form.

(2) Solve the linearized model. Let the value of the objective function be Z^1. Let the solution values of the decision variables be $\underline{X}^1$.

---

* Σ (Flow Along Plant Supply Paths) + (Flow Along Resupply Paths)

$\leq$ (Plant Capacity).

Note that constraint (c) in (4.18) requiring flow into a warehouse to equal flow out of the warehouse is met automatically by the transportation matrix [see 12].

(3) Calculate a new set of warehouse throughputs based on $\underline{X}^1$. Let this set be $\underline{D}^1$.

Nth Stage

(4) Transform the nonlinear model into a new linear form based on $\underline{D}^{n-1}$.

(5) Solve the linearized model. Let the value of the objective function be Z^n, and the solution set be $\underline{X}^n$.

(6) Compare Z^n and Z^{n-1}. If Z^n is less than or equal to Z^{n-1}, proceed. If not, terminate.

(7) Compare $\underline{X}^n$ and $\underline{X}^{n-1}$. If $\underline{X}^n$ is not equivalent (variable by variable) to $\underline{X}^{n-1}$, proceed. If it is, terminate.

(8) Use $\underline{X}^n$ to compute a new set $\underline{D}^n$.

(9) Go to step (4).

Note that there are two stopping rules employed in this routine. First, each successive Z value must be less than or equal to the previous value. If not, the procedure terminates. The purpose for this rule is two-fold. Cycling is prevented, and each step or iteration of the procedure is ensured to produce solutions at least as good as previous solutions.

The second stopping rule is that the procedure should terminate when two consecutive solutions repeat themselves. When this occurs, the values of the current throughput set ($\underline{D}$) and the previous throughput set are equivalent since all decision variables are equivalent. Again, there is a two-fold purpose. First, as mentioned in the second chapter, it is not difficult to envision consecutive solutions having identical Z values but different values of the decision variables. That is, comparison of Z values alone is not a sufficient termination criteria to stop

SINKS / SOURCES	CUSTOMERS $k, \hat{m}=1,A\ 1,B---NK,\ NM$	WAREHOUSES (Leased) $j, \hat{m}=1,A\ 1,B---NJ^P,\ NM$	(Public)	SLACK	CAPACITY
Warehouses:					
($j \in J^L$) $j, m=1,A$	W^L_{jkm} for $m=\hat{m}$	0		M	S
$1,B$		.		M	S
.	M for $m \neq \hat{m}$	.	M	.	.
NJ^L, NM		.		.	.
($j \in J^P$) $j, m=1,A$		.		.	.
$1,B$	W^P_{jkm} for $m=\hat{m}$	M	.	.	.
.	M for $m \neq \hat{m}$		.	.	.
NJ^P, NM			0	M	S
Plants:					
$i, m=1,A$				0	$CAP_{1,A}$
$1,B$	r_{ikm} for $m=\hat{m}$	S^L_{ijm} for $m=\hat{m}$	S^P_{ijm} for $m=\hat{m}$	.	.
.	M for $m \neq \hat{m}$	M for $m \neq \hat{m}$	M for $m \neq \hat{m}$	.	.
NI, NM				0	$CAP_{NI,NM}$
Demand: $d_{km}=$	$d_{1A}\ d_{1B}---d_{NK,NM}$	S S - - - - - - - - - - - - -	S	$\bar{S}$	

where $W^L_{jkm} = (A^L_{jkm} + \frac{B^L_{jm}}{(D_{jm})^{\frac{1}{2}}}$, $\qquad W^P_{jkm} = (A^P_{jkm} + \frac{B^P_{jm}}{(D_{jm})^{\frac{1}{2}}} + \frac{FC_j}{D_j})$,

M = large number, $\qquad S = \sum_k \sum_m d_{km}$, $\qquad \bar{S} = \sum_i \sum_m CAP_{im} - \sum_k \sum_m d_{km}$.

Figure 15. Transportation Matrix.

the routine upon repetition of a solution (which implies (4.18) is equivalent to (3.26).

The other purpose of this second stopping rule is to ensure that a least-cost solution to the original, nonlinear model (3.26) has been found. Since this termination criteria provides two consecutive throughput sets, $\underline{D}$, which are equivalent, then the solution to the last linearized model is also a solution to the nonlinear model. Further, since the last linearized solution cannot be improved upon, the nonlinear solution also cannot be improved upon--within the local area where the linear function is equivalent to the nonlinear function, that is, within the local area of this extreme point solution. To see this, note the following facts.

a) As long as $\sum_k X_{jkm} = D_{jm}$ for all j, m, the value of the original objective function (3.26) is equal to the value of the linearized objective function --regardless of the individual X_{jkm} values.*

---

* For example, let

$$Z = 2x_1 + x_2 + (x_1 + x_2)^{\frac{1}{2}}, \text{ and } Z^L = 2x_1 + x_2 + \frac{(x_1+x_2)}{(x_1 + x_2)^{\frac{1}{2}}},$$

so that

$$Z^L = \text{linearized function},$$

$$Z^L = (2 + \frac{1}{D^{\frac{1}{2}}})x_1 + (1 + \frac{1}{D^{\frac{1}{2}}})x_2 ,$$

$$\text{where } D = (x_1 + x_2).$$

Let D = 16 and a) $x_1 = 6$, $x_2 = 10$, then

$$Z = 12 + 10 + 4 = 26, \text{ and } Z^L = (\tfrac{9}{4})(6) + (\tfrac{5}{4})(10) = 26,$$ (cont'd)

b) For any change in X_{jkm} values which does not change D_{jm} for all j, m, the new value of the original function and of the linearized function will be the same. That is, there will be an equivalent change in both functions.

c) At termination of the solution routine, the results of the linear model cannot be improved upon. Also, $D^T_{jm} = \sum_k X_{jkm}$ for all j, m.**

d) Since there are no X_{jkm} values which can be changed in terminal solution to improve the value of the linearized function, there are also no X_{jkm} values which can be changed to improve the original function (so long as $\sum_k X_{jkm}$ remain equal to D_{jm} for all j, m).

e) Hence, in the local area of the extreme point associated with the throughput set of D_{jm} values, the terminal solution of (4.18) is a locally optimal solution of (3.26).

To summarize, the linearization-solution routine approximates the nonlinear problem with a linear form and solves this linearized model in a successive manner until either a higher-cost solution is found or until a solution locally optimal to the original problem is found.

---

* or b) $x_1 = 7$, $x_2 = 9$, then

$$Z = 14 + 9 + 4 = 27, \text{ and } Z^L = (\tfrac{9}{4})(7) + (\tfrac{5}{4})(9) = 27.$$

**The results hold only if the terminal rule utilized is $\underline{X}^T = \underline{X}^{T-1}$.

Initial Starting Point

This linearization-solution routine is the basis of Phase One. In the first step of Phase One, an initial starting point is generated as input to the linearization-solution routine, which is then used to determine one possible solution to the problem. There are several strategies which may be used to formulate this initial starting point. The approach used in the solution procedure is to base the starting throughput set on the a priori "favorable" throughput level of each warehouse.

To arrive at these levels, first note that there are two basic alternative supply paths which can be used to meet a customer's demand. Supply paths with a plant as a source can be used, or supply paths with a warehouse as a source can be used. Ordinarily, the plant-related supply path is the least-cost alternative since it does not include warehouse operating cost. However, due to the service time constraint and to the fact that consolidation of customer shipments can reduce transportation costs, some warehouse supply paths may be required in the optimal system.

Since warehouse operating costs are concave (see figures 6 and 7), consolidating warehouse throughputs should be profitable. If those warehouses for which consolidation would prove most profitable were known, then the model solution should be geared to favoring this set of warehouses. In other words, since the optimal distribution system will likely require use of some warehouses, and since consolidation of warehouse throughputs is profitable, the initial starting point should favor the set of warehouses for which a large assignment of customer demand would be profitable.

The two alternative supply paths for a customer are given in figure 16.

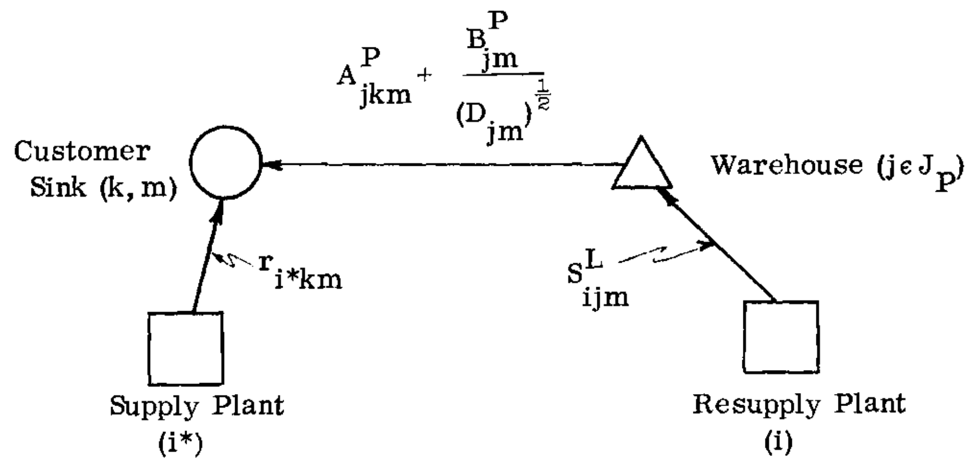


Figure 16. Alternative Supply Paths.

Based on the possibility of these two supply alternatives, a warehouse's "favorable" throughput should be derived from those customer-sinks for which the warehouse-related supply path costs (with operating costs evaluated at a large, consolidated level) are less than the plant-related supply path costs. All customers for which this cost relation holds are said to be members of that warehouse's "customer product set," $\underline{K}_{jm}$. Note that there will be a "favorable" throughput for each product. The sum of the demand of all elements of a warehouse's customer-product set is used as that warehouse's initial throughput for the associated product, m. That is,

$$D^{(Initial)}_{jm} = \sum_{k \epsilon \underline{K}_{jm}} d_{km}. \tag{4.19}$$

Note that a solution based on an initial starting point calculated in this manner will tend to favor those warehouses for which a consolidation of customer demand may prove profitable.

To summarize the formulation of this initial starting point, consider the following outline of the procedure.

(A) For each warehouse $j \epsilon J_p$, and each product m,

1) Determine those customers for which plant-to customer transportation costs are greater than the sum of plant-to warehouse transportation, warehouse-to-customer transportation, and warehouse operating costs. That is, all k for which

$$r_{i*km} > S^P_{\bar{i}jm} + A^p_{jkm} + \frac{B^P_{jm}}{(D)^{\frac{1}{2}}}, \tag{4.20}$$

where $i* = \left\{ i \mid r_{i*km} = \underset{i}{\text{minimum}} \; (r_{ikm}) \right\}$,

$$\bar{i} = \left\{ i \mid S^P_{ijm} = \text{minimum} \; (S^P_{ijm}) \right\},$$

$$D = \sum_k \sum_m d_{km}.$$

2) Let

$$\underline{K}_{jm} = \left\{ k \mid (r_{\bar{i}*km} > S^P_{\bar{i}jm} + A^P_{jkm} + \frac{B^P_{jm}}{(D)^{\frac{1}{2}}} \right\}. \tag{4.21}$$

3) Let

$$D^{(\text{Initial})}_{jm} = \sum_{k \epsilon \underline{K}_{jm}} d_{km}. \tag{4.22}$$

(B) For each warehouse $j \epsilon J_L$, and each product m,

1) Let

$$\underline{K}_{jm} = \left\{ k \,\middle|\, \left[r_{i^*mk} > S^L_{i\acute{j}m} + A^L_{ikm} + \frac{B^L_{jm}}{(D)^{\frac{1}{2}}} + \frac{FC_j}{D} \right] \right\}. \tag{4.23}$$

2) Let

$$D^{(Initial)}_{jm} = \sum_{k \epsilon \underline{K}_j} d_{km}. \tag{4.24}$$

Additional Starting Points

In order to eliminate the possibility of a suboptimal solution due to the intersection of two or more cost functions, additional successive linearization-solution routines are needed. The application of this routine using the initial starting point just described should lead to a "good" solution. However, as pointed out in the beginning of this chapter, a suboptimal solution can result when the linearization-solution routine (this routine is comparable to the "ordinary" successive linearization procedure of Hammond and others) is based on only one starting point. Recall that in order to ensure that suboptimality does not result from "curve crossing," the entire range of possible throughput volumes must be examined (that is, must be used in the linearization procedure). This may not be accomplished in a single application of the linearization-solution routine.

However, by using a series of additional starting points which cover the entire range of throughput values this suboptimality can be avoided. This can be accomplished by breaking the range of throughput levels (note that this range is from 0 to

$\sum_k \sum_m d_{km}$) into a finite number of intervals and basing a separate linearization-solution routine on each interval. That is, the beginning of each interval serves as a new starting point for an additional linearization solution routine.

Use of these additional starting points also serves another purpose--it provides explicit consideration to additional locally optimal extreme points. In this manner, the possibility is reduced that the final model solution will result in a local optimum rather than a global one. That is, these additional linearization-solution routines provide additional searches of the problem's solution space for optimal extreme points.

To see this, note that the linearization-solution routine starts at some extreme point or vertex of the original problem's convex feasible solution space, C, and progressively examines other extreme points until one is visited twice (consecutively). At each visit of an extreme point a linear surface is used to approximate or represent the original nonlinear surface (objective function). This linearized surface is based on the warehouse throughput set corresponding to that extreme point.

Based on this linearized objective function, C is examined for extreme points which offer a reduced value of this linear function. That is, the linear problem is solved. When an extreme point offering improvement can be found, the search of C moves to this new point, a new linear function is formed, and C is again searched for improvement points.

By requiring $Z^n \leq Z^{n-1}$ before moving on to the next extreme point (that is, stopping if $Z^n > Z^{n-1}$) the routine ensures that cycling will not occur. Also, the solution corresponding to each extreme point explicitly visited must be at least no worse than previous ones. Once an extreme point is examined twice (consecutively), a local optimum to the original nonlinear problem has been found. The linearization-solution routine for this one search of C then terminates. Note that this process has both implicitly and explicitly enumerated extreme points for consideration. Figure 17 gives a graphic explanation of the search of C for a solution.

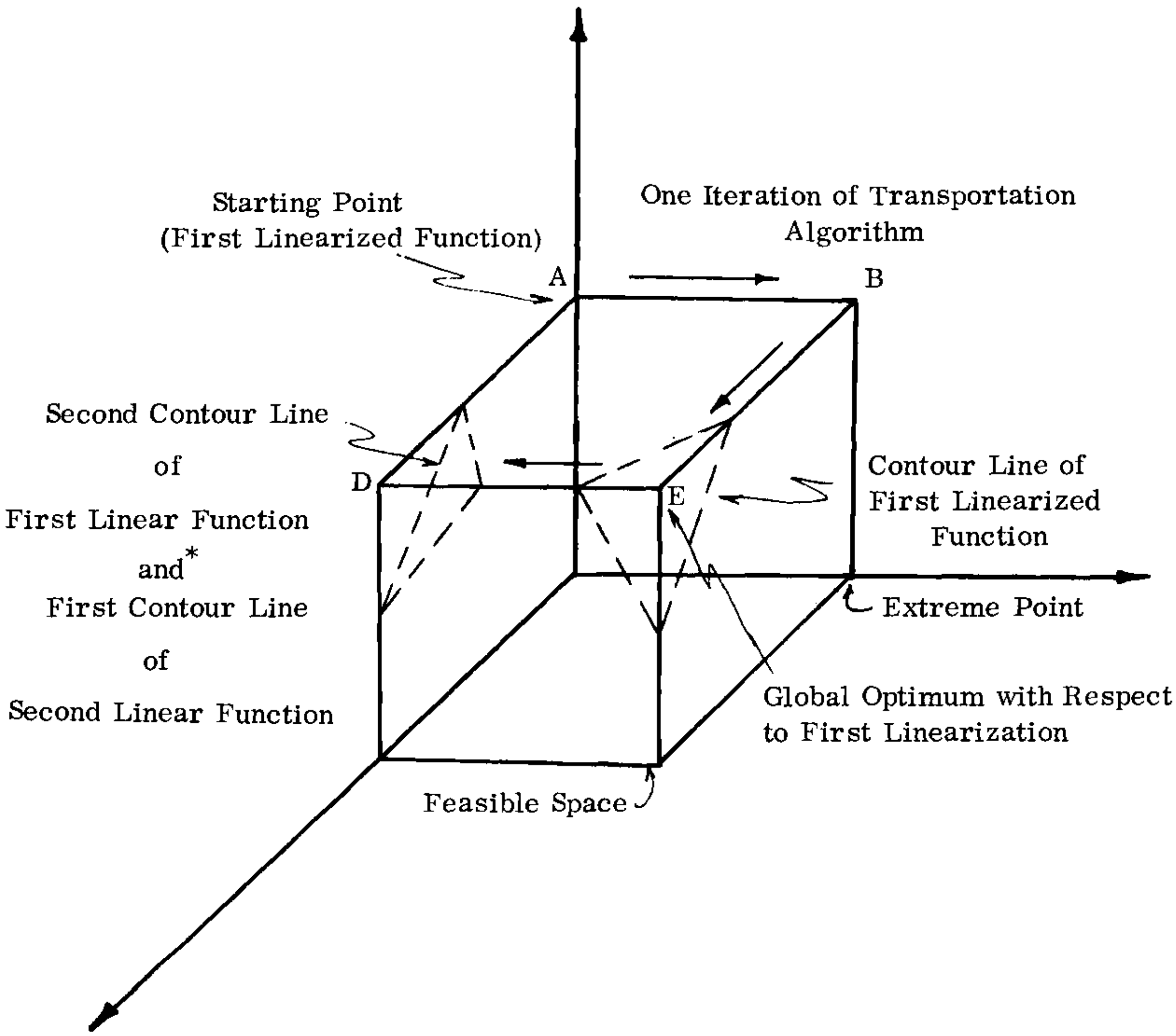
Phase One contains several such search procedures of the solution space C. Each search begins at a different starting point. An entirely different set of extreme points may be visited and enumerated on each separate search, or two "trips" through C may repeat themselves. The end results of each search is an extreme point which is a local optimum of the original nonlinear problem.* Since the final solution of each pass through C may be different, that solution corresponding to the least-cost result should be used as the end-product of the first phase. This solution forms the starting point of Phase Two. Figure 18 summarizes the entire first phase by giving a flow diagram of the steps involved (note that three starting points in addition to the initial point are used; that is, I = 3).

Phase Two

Two objectives of the multi-search procedure of Phase One are the elimination of a possible suboptimal solution due to "curve-crossing," and the examination

---

* If the linearization-solution routine terminates due to $Z^n > Z^{n-1}$, it can not presently be shown that a local optimum has (or has not) been found. Termination due to this criteria is not thought to occur often.



*That is, use of extreme point D to linearize the cost function does not change the plane (that is, $\underline{X}^2 = \underline{X}^3$); hence the procedure will terminate with D as solution.

Figure 17. Search of Solution Space, C.

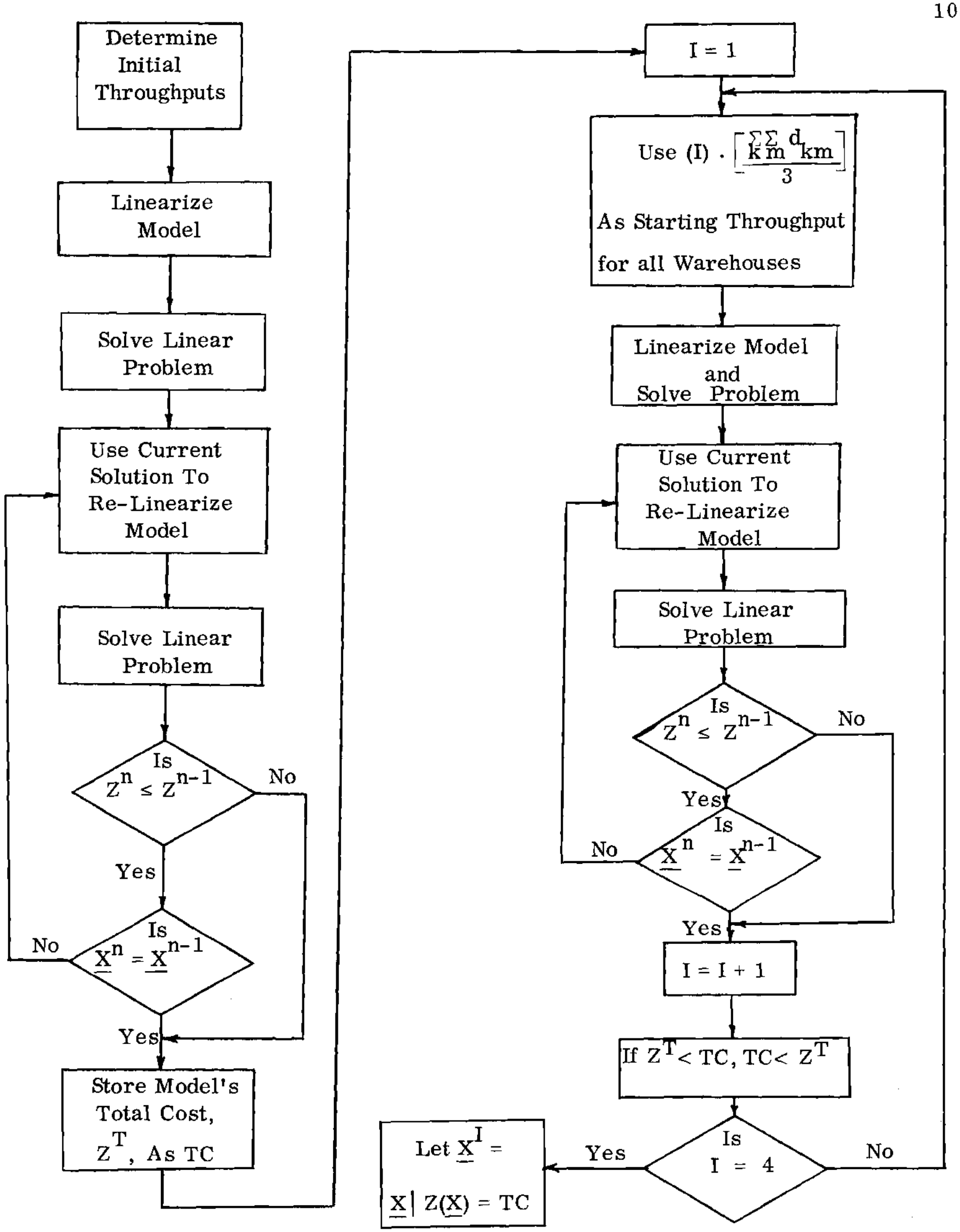


Figure 18. Flow Diagram of Phase One Procedure.

of more extreme points for optimality. The objective of Phase Two is to deal with the other theoretical consideration mentioned earlier--the fact that successive linearization may tend to bias its results. Recall that this bias is the tendency for solutions to specify the use of more warehouses than is optimal.

The second phase of the solution procedure is based on the assumption that the results of Phase One do in fact exhibit this bias. That is, Phase Two attempts to improve the solution found in the first phase by "encouraging" the consolidation of warehouse usage.

The "encouragement" takes the form of a penalty cost and is attached to all warehouses <u>other than</u> those for which an increase in size (throughput level) would seem to reduce total system costs. In other words, a penalty cost is added to the unit operating costs of all warehouses not identified as profitable candidates for reassignment of customer demand. In effect, this penalty cost encourages the dropping of warehouses from consideration in a similar manner to the "drop approach" used in the heuristic location algorithm of Feldman, Lehren, and Ray [8].

The decision process of the solution analysis is concerned with determining the source (warehouse or plant) to which a customer-sink's demand should be assigned. Therefore, one approach to <u>formulating</u> the required penalty <u>functions</u> is to determine the <u>one</u> most "favorable" warehouse candidate for consolidation as related to <u>each</u> separate customer-sink. All warehouses except a customer-sink's "most favorable" warehouse are penalized for shipments from <u>that</u> sink. Note that it is possible for a warehouse to be assigned a penalty cost for one customer-sink's assignment but not for another's assignment.

A warehouse is "favorable" with respect to a customer-sink if, when operating at a large, consolidated throughput level, that warehouse forms a lower-cost supply path than the least-cost plant-related supply path. In terms of the "customer-product set," $\underline{K}_{jm}$, defined in the discussion of Phase One, a warehouse j is "favorable" for a customer-sink if the associated customer is in the set $\underline{K}_{jm}$. Note that this definition is based on the problem characteristic of allowing "direct" shipments (the initial starting point of Phase One also utilizes this characteristic).

Since a customer will probably be included in more than one customer-product set, that customer will have more than one "favorable" warehouse. The "favorable" warehouse which is "favorable" to the most sinks will obviously be the best candidate for consolidation with respect to supplying <u>those</u> sink demands. Hence, a customer-sink's "most favored," or simply "favored," warehouse can be defined as that sink's "favorable" warehouse j having the largest (demand-wise) customer-product set $\underline{K}_{jm}$. For customer-sink k-m, call this "favored" warehouse $(\bar{j})^{km}$.

This formulation of the penalty function will tend to identify and "favor" a profitable warehouse in several geographical areas. That is, customers in the same geographical region will tend to "favor" the same warehouse, while a different warehouse will be "favored" by customers in some other region. The results of <u>several</u> warehouses across the country which appear to be the most profitable choices for the <u>one</u> warehouse to utilize <u>in each region.</u>

There are several possible ways to derive the expression of the unit penalty cost attached to a customer-sink's "unfavorable" warehouses. The approach taken in Phase Two is that this penalty should reflect the average unit cost of not assigning a customer to his "favored" warehouse. That is, the penalty cost is an "opportunity" cost. It could be thought of as a negative penalty assigned to a customer's "favored" warehouse.

The first step in calculating the (negative) penalty cost associated with a "favored" warehouse, $(\bar{j})^{km}$ to determine the total transportation plus warehouse operating costs if $(\bar{j})^{km}$ is used alone in the region in which it is a "favored" warehouse. That is, determine the total cost of using only $(\bar{j})^{km}$ in its region (its region is the area covered by the network of customers in j's customer-product set $\underline{K}_{jm}$).

The next step is to determine the total costs resulting from the Phase One solution associated with those sinks in the $(\bar{j})^{km}$ region who are potential assignments to $(\bar{j})^{km}$ (that is, all $k\text{-}m/k \epsilon \underline{K}_{jm}$). The difference in these two total costs figures is the "opportunity" cost of not using $(\bar{j})^{km}$ in the Phase One solution. By dividing this difference by the sum of the demand of all sinks considered in this difference (that is, all $k\text{-}m/k \epsilon \underline{K}_{jm}$), an average unit penalty cost can be had.

As a summary of Phase Two, consider the following list of steps in this Phase.

1. Determine each customer-sink's set of "favorable" warehouses, $\underline{J}_{km}$. This is the set of those warehouses (j) for which the plant-to-warehouse transportation cost plus the warehouse operating cost (evaluated at a large, consolidated operating level) is less than the supply path costs from the closest

supply plant to that customer sink. That is, for each k, m, let

$$\underline{J}_{km} = \begin{cases} \left\{ j \in J_P \mid \left[S^P_{\bar{i}jm} + A^P_{jkm} + \dfrac{B^P_{jm}}{(D)^{\frac{1}{2}}} \right] < r_{i^*km} \right\}, \\ \left\{ j \in J_L \mid \left[S^L_{i'jm} + A^L_{jkm} + \dfrac{B^L_{jm}}{(D)^{\frac{1}{2}}} + \dfrac{FC_j}{D} \right] < r_{i^*km} \right\} \end{cases} \tag{4.25}$$

$$\text{where} \quad i^* = \left\{ i \mid r_{i^*km} = \underset{i}{\text{minimum}} \, (r_{ikm}) \right\},$$

$$\bar{i} = \left\{ i \mid S^P_{\bar{i}jm} = \underset{i}{\text{minimum}} \, (S^P_{ijm}) \right\},$$

$$i' = \left\{ i \mid S^L_{i'jm} = \underset{i}{\text{minimum}} \, (S^L_{ijm}) \right\},$$

$$D = \sum_k \sum_m d_{km}.$$

Note that an alternative criteria to use in determining $\underline{J}_{km}$ would be to allow all warehouses which are associated with feasible supply paths for customer-sink k-m to be elements of $\underline{J}_{km}$. That is, for sink k-m the set $\underline{J}_{km}$ would include all those j for which the j to k-m transportation delivery time is less than or equal to the service limit time constraint. This would result in a larger set of "favorable" warehouses, $\underline{J}_{km}$.

2. Determine each warehouse's "customer-product set," $\underline{K}_{jm}$. For warehouse j, and product m, this is the set of all customers for which j is a number of the associated customer-sink's "favorable" warehouse set. That is, for

each j, m, let

$$\underline{K}_{jm} = \left\{ k \mid j \epsilon \underline{J}_{km} \right\} . \tag{4.26}$$

Note that this is equivalent to the definition of $\underline{K}_{jm}$ given previously in (4.21) and (4.23).

3. Determine the total demand associated with each warehouse's "customer-product set." Let this cumulative throughput level be D^{II}_{jm}. That is, for each j, m, let

$$D^{II}_{jm} = \sum_{k \epsilon \underline{K}_{jm}} d_{km} . \tag{4.27}$$

4. For each customer-sink determine that member of its "favorable" warehouse set which is the "most favored," that is, which has the largest D^{II}_{jm}. Designate this warehouse as $(\bar{j})^{km}$. That is, for each k, m, let

$$(\bar{j})^{km} = \left\{ j \mid D^{II}_{\bar{j}m} = \underset{j \epsilon \underline{J}_{km}}{\text{maximum}} \, (D^{II}_{jm}) \right\} \tag{4.28}$$

5. Formulate and assign a unit penalty cost, P_{jkm}, to each warehouse-related supply path for customer sink k-m. This penalty cost is the average opportunity cost associated with not assigning k-m to its "favored" warehouse, $(\bar{j})^{km}$. That is, for each k, m and $j = (\bar{j})^{km}$, let $P_{jkm} = A/B$,

where $A = \Big[$(Sum of warehouse operating cost associated with those customers in set $\underline{K}_{jm}$ in the final Phase One solution) + (sum of Phase One transportation costs of supply and resupply paths for $k \in \underline{K}_{jm}$)--(total operating plus transportation costs if j is used alone)$\Big]$,

$B = \Big[$Sum of the annual demand of all customer members of $\underline{K}_{jm}\Big]$.

In other words, for each k, m and $(\bar{j})^{km} \in J_P$,

$$P_{jkm} = \frac{A^P}{B},$$

$$\text{where } A^P = \Bigg\{ \sum_{k \in \underline{K}_j} \sum_m \Bigg[\Bigg[A^P_{jkm} + \frac{B^P_{jm}}{(D^I_{jm})^{\frac{1}{2}}} \Bigg] X^I_{jkm} + \Big[r_{i*km} X^I_{i*km} + \sum_i S^P_{ijm} X^I_{ijm} \Big] \Bigg] - \sum_{k \in \underline{K}_j} \sum_m \Bigg[\Bigg[A^P_{jkm} + \frac{B^P_{jm}}{(D^{II}_{jm})^{\frac{1}{2}}} \Bigg] D^{II}_{jm} + S^P_{ijm} D^{II}_{jm} \Bigg] \Bigg\},$$

$$B = \sum_{k \in \underline{K}_j} \sum_m d_{km},$$

$$\text{where } D^{II}_{jm} = \sum_{k \in \underline{K}_j} d_{km},$$

II $\Rightarrow$ the value of the respective parameter or variable associated with the initial starting point of Phase Two,

I $\Rightarrow$ the value of the respective parameter or variable associated with the final solution of Phase One.

In a similar manner, P_{jkm} for $(\bar{j})^{km} \epsilon J_L$ could also be defined.

6. Let the unit cost in the initial linearized model of Phase Two be those unit cost in the final solution of Phase One plus the penalty cost where appropriate. That is, for each k, m let

$$C^{II}_{jkm} = C^{I}_{jkm} + P_{jkm}, \text{ for all } j \neq (\bar{j})^{km}, \tag{4.30}$$

and

$$C^{II}_{jkm} = C^{I}_{jkm}, \text{ for } j = (\bar{j})^{km},$$

where C^{I}_{jkm} = final Phase One unit cost associated with supply path j-k-m,

C^{II}_{jkm} = initial Phase Two unit cost for path j-k-m.

7. Leaving all plant-related supply path costs, and all resupply path costs unperturbed, use the linear costs computed in step (6) as cost coefficients parameters in the initial linearized problem for Phase Two.

8. Successively iterate the linearization-solution routine described previously until the stopping criteria is met.

9. The best of the Phase One and Phase Two results is the final solution of the original model (3.26).

Justification of Solution Procedure

Model (3.26) could have been attacked by several alternative solution approaches. As mentioned in the second chapter, mathematical programming in the form of a zero-one integer program could have been used. To formulate

(3.26) in this mode would require that all discontinuous, strictly concave warehouse operating cost functions be approximated by a series of linear segments. A zero-one variable would be associated with each segment for each product for each warehouse. For a model containing 40 warehouses (each with the option of being public or leased), four product lines, and five break points in each warehouse operating curve, at least (40)·(2)·(4)·(5) = 1600 zero-one variables would be required in the mixed integer formulation. The computational and storage requirements of this formulation would be extreme. However, the results of this approach would be an exact solution to the approximated model.

On the other hand, the two-phase successive linearization approach described above offers a "good" approximate solution to the exact model. This solution approach can be justified on two bases--the solution procedure "efficiently" results in local optimum, and is an improvement over similar approaches advocated in the literature.

The two-phase solution procedure can be characterized as an "efficient" and effective search procedure. It is "efficient" in the sense that extreme point solutions are examined not only explicitly but also implicitly. This occurs in the linearization-solution routine in a manner equivalent to implicit enumeration of solution points by the well-known Simplex procedure. The solution procedure is "effective" since it results in a locally optimal solution.

As mentioned in the discussion of the linearization-solution routine, a local optimum resulting as a solution is seen in the fact that upon termination, the solution to the linearized model (4.18) is equivalent to a solution of the original model (3.26). This equivalence was shown in the previous section on Phase One. Due to the fact that

a) changes in X_{jkm} values which do not disturb the terminal relations $\sum_k X_{jkm} = D_{jm}$ for all j, m, cause equivalent changes in <u>both</u> the linearized function and the nonlinear function, and*

b) the solution of the linearized model (4.18) at termination cannot be improved upon in any area,

then the value of the original model cannot be improved upon in a local area. That is, the Phase One (and Phase Two) solution is a local optimum.

Baumol and Wolfe [10] stated that their successive linearization solution procedure resulted in local optimum. Even though they did not elaborate on this conclusion, their contention is backed by both Balinski [2, p. 286] and Kuehn and Hamburger [7, p. 661].** The Baumol and Wolfe nonlinear model is basically equivalent to the nonlinear model (3.26). Also, the linearized form used by Baumol and Wolfe to successively solve the original problem is equivalent to the current linearized model (4.18) except for a constant multiplier which can be disregarded. Further, Baumol and Wolfe's successive linearization

---

* Recall that this is true only if the stopping criteria of $\underline{X}^T = \underline{X}^{T-1}$ is utilized.

** The Kuehn and Hamburger definition of a local optimum to the location problem is that "no individual unit of product can be shipped by an alternative route without increasing total distribution cost." Note that this does not involve the location or number of warehouses, but only the size of warehouses.

routine is comparable to one complete iteration of the linearization-solution routine mentioned above. Hence, if the Baumol and Wolfe solution is a local optimum, it follows that the results of each iteration of the two-phase solution procedure is also a local optimum.

A second justification of the two-phase solution procedure is that it is an extension of similar successive linearization approaches reported in the literature. This reasoning is not to imply that successive linearization is the most effective or efficient solution procedure available for solving this type problem. Rather, successive linearization is a standard solution approach advocated in the literature (standard implying benchmark). By modifying and extending this standard solution method for the purpose of improving the results, the solution of the two-phase procedure should meet or surpass standards set for solution methods for the original model (3.26). Modifications incorporated into the two-phase procedure include in Phase One an improved starting point, a multi-search of the solution space, and insurance against suboptimization due to "curve crossing." The additional Phase Two routine is an attempt to recognize and select a solution which is an improvement over the local optimum found in Phase One.

CHAPTER V

APPLICATION

In order to test the method of solution of model (3.26), an example problem was synthesized and solved. This problem can be described as follows:

Example Problem

Given:

(1) two production facilities with fixed locations and unlimited capacities,

(2) two alternative transportation modes available for resupply paths,

(3) a nation-wide demand for a single product,

(4) a three-day service time limit, and

(5) two alternative types of warehousing, leased and public,

Find:

(a) the number, locations, and sizes of warehouses to use,

(b) whether each warehouse used should be of a leased or public type,

(c) the production facility to use to resupply each warehouse,

(d) the resupply mode,

(e) the supply source to use to deliver each customer's demand,

Subject to:

(I) all customer demand must be satisfied,

(II) supply path delivery times must be less than or equal to three days.

Note that this problem can adequately be represented by model (3.26) by disregarding the plant capacity constraints. Also, since this is a single-product (or a single product-line) problem, there will be only one product multiplier.

As this is a multi-customer, nation-wide distribution system, the concepts of warehouse location zones and customer zones should be employed. One feasible zoning structure is given in figure 19. The size, shape, and number of zones and the choice of key cities is based on an attempt to force this example problem to resemble the problem posed by Baumol and Wolfe. Their problem contained two fixed plant sites (with limited capacities), five fixed warehouse sites, and eight fixed customer demand centers (with no provision for "direct" shipments).

Data Base

The data used in this example problem is basically the data used in the Baumol and Wolfe problem. There are four categories of data used in the model: customer demand, resupply tariff data, supply tariff data, and warehouse operating costs. This data can be represented in Tables 1-4. (Recall that the index i represents plants, k stands for customers, j for warehouses, and w for transportation modes).

There are two volume break points at which handling and storage rates in public warehousing will be reduced (due to renegotiated contracts). These are 30 units and 60 units. Handling rates will be reduced by .10 dollars at each break point, and storage rates reduced by one dollar. Likewise, leasing fees will be

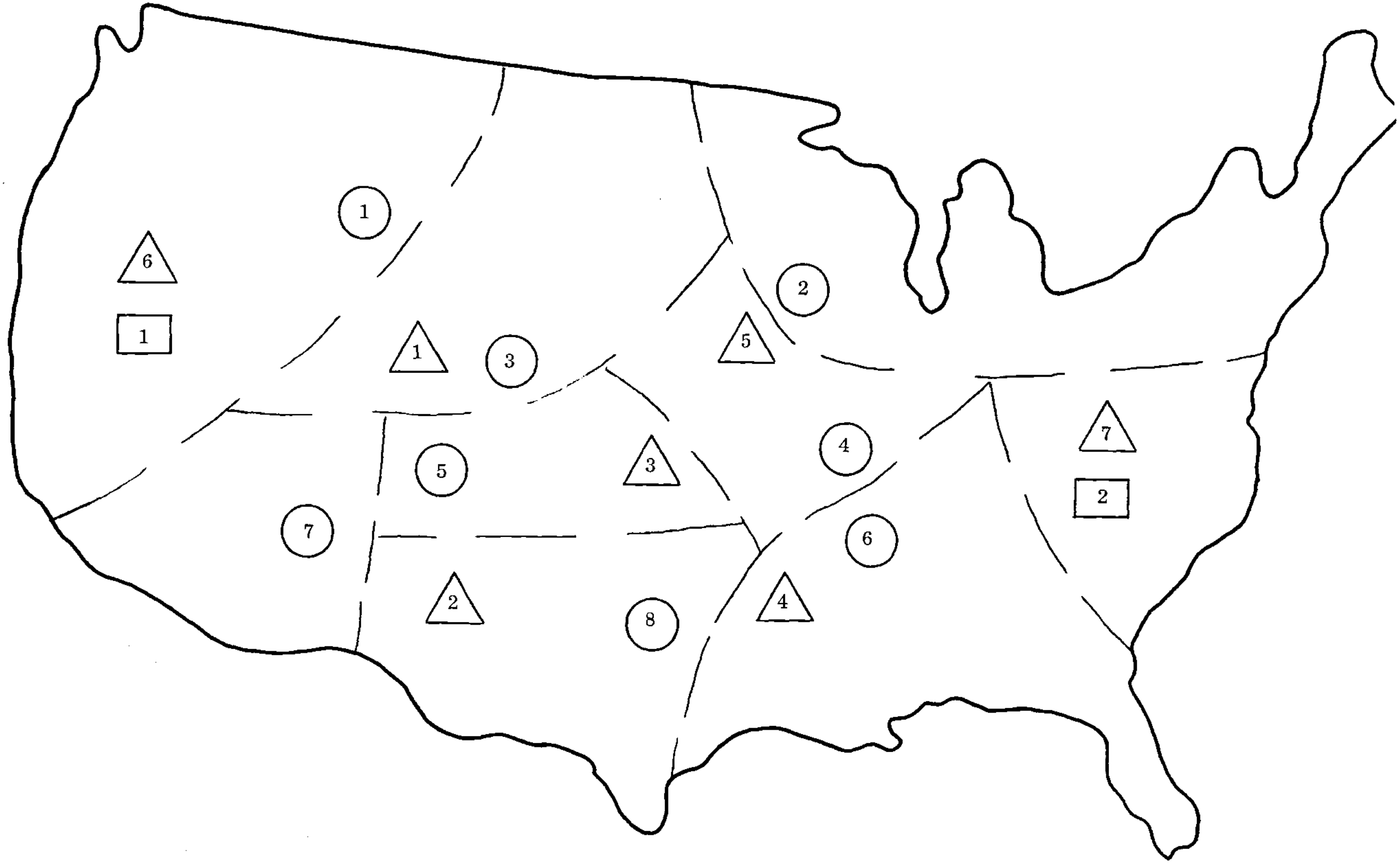


Figure 19. Warehouse Location Zones and Customer Zones Used in Example Problem.

Table 1. Customer Demand (Units)

Month =	1	2	3	4	5	6	7	8	9	10	11	12	Total
k=1	0	0	1	2	5	2	0	0	0	0	0	0	10
2	0	1	1	1	2	2	2	1	0	0	0	0	10
3	1	1	1	1	1	1	1	1	1	1	0	0	10
4	1	2	1	2	4	2	1	1	1	0	0	0	15
5	0	0	0	1	2	2	0	0	0	0	0	0	15
6	0	1	2	1	2	5	2	1	1	0	0	0	15
7	1	1	1	1	1	1	1	1	1	1	0	0	10
8	0	0	0	1	3	4	7	0	0	0	0	0	15

Table 2. Resupply Tariff Data (Rates in dollars per unit)

		w=1		w=2	
		LT	Tariff Rate	LT	Tariff Rate
i=1	j= 1	7	7	10	5
	2	12	7	15	5
	3	7	8	8	4
	4	15	12	18	9
	5	10	11	14	9
	6	1	2	1	2
	7	21	20	21	20
i=2	1	20	14	21	10
	2	15	12	18	10
	3	10	9	14	6
	4	5	6	7	3
	5	7	8	10	6
	6	21	20	21	20
	7	1	2	1	2

Table 3. Supply Tariff Data (Rates in dollars per unit)

k =	1		2		3		4		5		6		7		8	
	LT	R	LT	R	LT	R	LT	R	LT	R	LT	R	LT	R	LT	R
i=1	<*	25	>	99**	<	22	>	99	<	22	>	99	<	21	>	99
2	>	99	<	26	>	99	<	21	>	99	<	22	>	99	<	25
j=1	<	5	<	11	<	3	<	8	<	5	<	10	<	11	<	11
2	<	14	>	99	<	8	<	9	<	4	<	7	<	4	<	4
3	<	10	<	11	<	3	<	5	<	2	<	5	<	9	<	5
4	>	99	<	13	<	9	<	6	<	7	<	2	<	10	<	2
5	<	9	<	7	<	3	<	2	<	6	<	5	<	12	<	8
6	<	25	>	99	<	22	>	99	<	22	>	99	<	21	>	99
7	>	99	<	26		99	<	21	>	99	<	22	>	99	<	25

* < implies that the associated lead time is <u>greater than</u> the three day service limit. Likewise,

> implies a LT <u>less than</u> (or equal to) the limit,

** 99 is the large unit cost attached to infeasible supply paths.

Table 4. Warehouse Operating Costs

		j=1	2	3	4	5	6	7
Public	Handling charge (dollars per unit)	2	1	2	2	3	3	3
	Storage charge (dollars per unit)	5	6	6	5	7	5	5
Leased	Leasing fee (dollars per unit)	10	10	10	10	10	10	10
	Fixed Cost (dollars)	80	80	80	80	80	80	80

reduced by two dollars at 30 units and one dollar more at 60 units. However, due to the increased requirement for company-owned handling resources at high operating levels, fixed cost will increase by 70 dollars at both 30 and 60 units of throughput. Also, unit production cost at both plants is assumed to be one dollar.

Computer Code

The computer program written to solve this distribution problem (as well as solve the slightly more general problem represented by model (3.26)) was coded in Fortran IV. It is based on the two-phase solution procedure outlined in the fourth chapter.

In order to increase the efficiency of the program, three assumptions involving the underlying problem were made. While not essential to the solution procedure, these assumptions allow the problem size to be reduced substantially.

The first assumption involves unit production costs. Production costs are used in two places in the model. They are added to plant-to-customer and plant-to-warehouse transportation costs, and are used in determining inventory holding cost at each warehouse. The assumption being made is that the average production cost of all items produced at a plant can be used in lieu of actual production costs for individual items at that plant.

The purpose of this assumption, as well as the other two assumptions, is to remove the need for explicit consideration of the product index (m) in plant-to-warehouse paths. This allows the number of columns and rows in the embedded transportation matrix to be reduced. Since inventory costs are being included in warehouse-to-customer paths, the use of average costs in inventory cost calculations will not help reduce the problem size. Therefore, actual cost can be used in determining inventory holding costs--provided a complimentary second assumption is made.

This assumption is that the unit production cost associated with the "least cost" resupply plant for a warehouse can be used in calculating holding costs at that warehouse. A discussion of this assumption was given in the third chapter. Essentially, this assumption allows for elimination of the plant subscript in the warehouse-to-customer component of warehouse operating costs (see figure 15).

The third assumption is that the pattern of demand of each product over time is the same. That is, the forecast for all products includes the same cycles and trends, with similar magnitudes. Use of this assumption eliminates the need for calculating (and using) a separate safety stock and reorder point coefficient for each product. That is, $SS^{w}_{ij} \Rightarrow SS^{w}_{ijm}$, and $RP^{w}_{ij} \Rightarrow RP^{w}_{ijm}$, for each m.

By incorporating these three assumptions into a problem, significant reduction in the size of the problem can be achieved. For example, a model containing 43 warehouse zones, 64 customer zones, 4 products, 5 plants, and 2 alternative types of warehousing can be reduced from a problem with 600 sinks (see figure 15) and 350 sources to one of 350 sinks and 90 sources. Over 66,000 elements (or, 31 per cent) of the embedded transportation matrix are eliminated.

Obviously, storage requirements and computation time are greatly reduced.

Another feature incorporated into the computer code is the use of the previous basis as the starting basis of the embedded transportation algorithm.* That is, in the successive linearization-solution routine the basis corresponding to the last optimal linear solution is used as the initial basic feasible solution for each current solution procedure. For large problems, the efficiency of the linearization-solution routine segment of the program should be significantly increased.

Solution

The example problem was solved on a Burrough's 5500 computer using the code referred to above. Solution times were under three minutes.

In order to test the sensitivity of this particular problem to changes in the "product multiplier," two different multiplier values were tried. The problem was first solved with a value of one. The final solution for this run is given in figure 20. Note that warehouse 3 is the only warehouse utilized in the solution. It is of the "leased" type, which is reasonable in view of its relatively large throughput. Also, note that the second alternative transportation mode is the least-cost resupply mode for warehouse 3.

This solution resulted from Phase Two of the solution procedure. Phase One found a solution with a throughput set assigning 20 units to warehouse 1, 35 units to warehouse 4, and 25 units to warehouse 5. The total cost of this solution

---

*This algorithm is based on the primal method presented by Hadley [17].

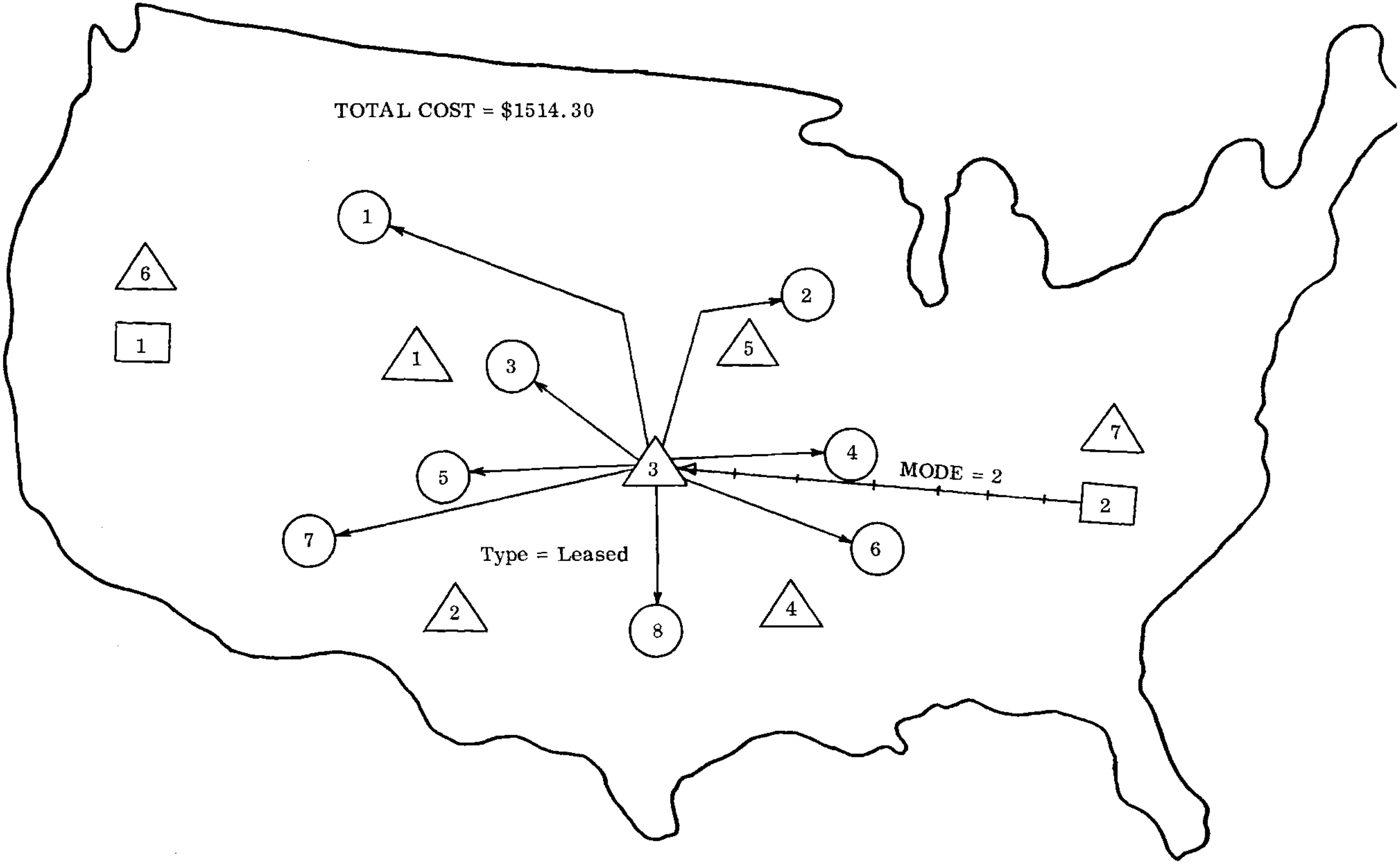


Figure 20. Example Problem Solution (with Product Multiplier Equal to One).

is $1869.00.

Phase Two identified warehouse 3 as the "most favorable" warehouse for all customers. The penalth costs formulated were sufficient to force the consolidated use of warehouse 3 by all sinks. The total cost of the second phase solution was $1514.30. Hence, the Phase Two results constitute the final solution to the problem--for a "product multiplier" of one.

A second run was made using an updated multiplier of two. Due to the larger inventory operating costs in this second formulation of the problem, the Phase One results utilized "direct shipments" to supply all customer demand, except for 30 units which were assigned to warehouse 4. The total cost of this solution is $2017.70.

However, as was the case in the first version of the problem, application of Phase Two resulted in a lower-cost solution. Again, warehouse 3 was identified as the "most favorable" warehouse for all customers. Based on the penalty costs attached to all appropriate source-sink paths, the second phase solution assigned all customers to warehouse 3, except customer 7 who was supplied "directly" from plant 1. The total cost of this solution is $1706.80.

This particular example problem proved to be sensitive to changes in the "product multiplier." The purpose of the sensitivity analysis of the multiplier value was not to establish the "correct" value to use (as stated previously, this is a single item problem implying a "correct" multiplier of one), but rather to indicate the sensitivity of the solution to changes in this value. The sensitivity found in this example indicates that an investigation should be conducted to determine the actual throughput of each individual item in the composite product (if the single product in the example in fact represented a composite of items). That is, the actual multiplier value corresponding to the initial solution should be established through an external investigation of which customers are included in the throughput of each warehouse. From this investigation, the actual multiplier can be determined which equates approximate and true inventory costs at each warehouse. This actual value would then be introduced as the multiplier and the problem re-solved.

CHAPTER VI

CONCLUSIONS

Research Conclusions

The research undertaken to formulate model (3.26) and develop the two-phase solution procedure presented in the fourth chapter led to several conclusions. These are included in the following listing.

Model

1. All possible location sites are either explicitly or implicitly considered by the model through the use of "warehouse location zones."
2. For cases of relatively homogeneous demand patterns among individual customers, the use of "customer zones" is an effective modeling tool. The size of the model as well as the data-gathering requirements can be substantially reduced by employing this zone concept.
3. A customer service or delivery time limit can be effectively incorporated into the solution analysis as an external data constraint.
4. Multi-product distribution systems present no difficulties when system costs are linear. However, when nonlinear costs are involved, the multi-product case cannot be modeled in a "straight-forward" manner. For example, if product lines are used to represent several individual products, "product multipliers" must be used in the model to correct for the understatement of inventory storage costs.

5. By making use of their definitions, safety stock and reorder point levels can be expressed as functions of annual warehouse throughput and included as model parameters. These functions can be included whether they are of a convex or nonconvex form.
6. There is a cost trade-off in the distribution system between resupply transportation costs and inventory storage costs. This trade-off is based on the choice of resupply delivery modes (and their corresponding lead times).

Solution Procedure

1. The solution procedure allows subjective as well as objective factors to be used in the selection of warehouse location sites.
2. The choice of supply and resupply modes can be made on the basis of procedures external to the solution of model (3.26).
3. The concept of a penalty function in the second phase of the solution procedure is, theoretically, an effective approach of forcing a move from a local to a global optimum. However, the specific formulation of the function is subject to question.
4. The characteristics of the distribution system underlying the problem at hand have a significant effect on the effectiveness of the two-phase solution procedure. For example, the homogeneity of production costs and of demand patterns affects the validity of several basic model assumptions. Also, the degree of concavity of the objective function affects the association between the nonlinear and the linearized models.

Solution

1. The two-phase successive linearization solution procedure results in a solution which is a local optimum to the original nonlinear problem.
2. The multi-search of the solution space (corresponding to the successive reapplication of the linearization-solution routine) does not necessarily result in the "best" local optimum.

Recommendations

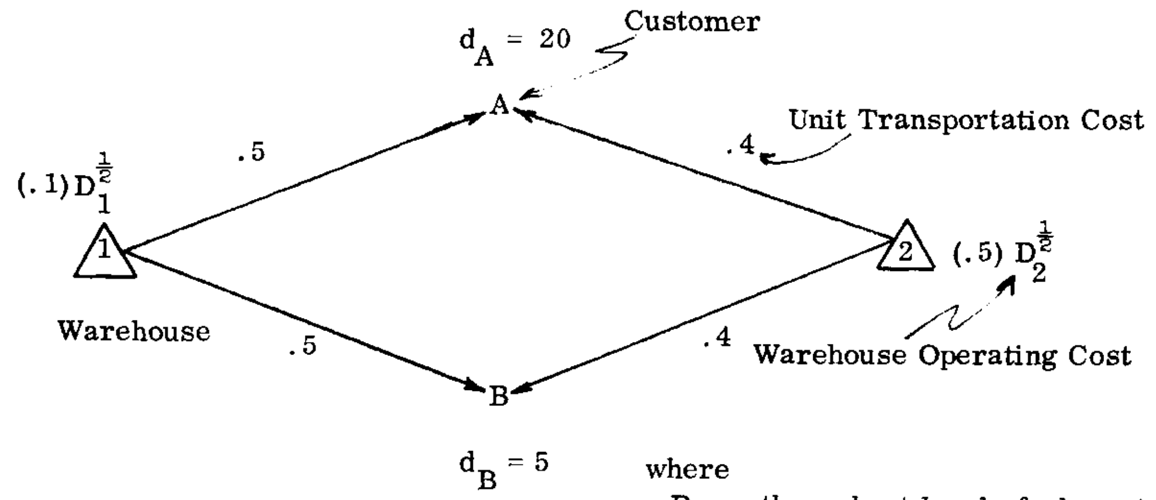
Several areas of further investigation which would extend or modify the current research are presented in the following list.

1. Incorporate the service limit restriction as an internal model constraint.
2. Develop the "best" closed form of a "product multiplier." This might eliminate the need for a post-solution sensitivity analysis.
3. Develop a more effective formulation of the Phase Two penalty functions.
4. Formulate a lower bound on the value of the nonlinear problem and incorporate this bound into the solution procedure.
5. Incorporate a simulation model as a third phase in the solution procedure. That is, let the final solution of the two-phase procedure constitute a starting point for a simulation study.
6. Develop a solution procedure which would necessarily lead to a global optimum. Since local optima are found when $D^T_{jm} = \sum_k X^T_{kjm}$, for all j, m, a branch and bound search of the solution space might be developed to enumerate these local optima.

7. A post-solution analysis should be developed to determine the best "mix" of the present physical distribution system and the theoretically optimum distribution system. This analysis might be based on should factors as the closing costs of eliminating current facilities, and subjective "accommodation" decisions.
8. Develop criteria to evaluate the decision of whether to formulate an exact model of a system and use approximate solution techniques, or to formulate an approximate model and determine an exact solution. This decision is applicable to problems such as the nonconvex distribution problem. As pointed out previously, the former model-solution approach was used in the current research, while the latter approach would be characteristic of an integer programming formulation of the problem.
9. Reformulate the model so as to develop a profit-maximization problem rather than a cost-minimization. This would allow "private" carriers to be considered as alternative transportation modes, as well as allow consideration to be given to the alternative of "owned" warehouse space.

APPENDIX

As an example of the linearization approach (Hammond's) resulting in a suboptimum solution due to the intersection or "crossing" of two or more cost functions, consider the following problem:



where

D_j = throughput level of whse. j

d_k = annual demand of customer k

1. <u>Model</u>

$$\text{MIN } Z = (.5X_{1A} + .5X_{1B} + .4X_{2A} + .4X_{2B}) + (.1)(X_{1A} + X_{1B})^{\frac{1}{2}} + (.5)(X_{2A} + X_{2B})^{\frac{1}{2}}$$

$$\text{ST: } X_{1A} + X_{2A} = 20$$

$$X_{1B} + X_{2B} = 5$$

or, linearizing the model,

$$\text{MIN } Z = (.5 + \frac{.1}{D_1^{\frac{1}{2}}})\, X_{1A} + (.5 + \frac{.1}{D_1^{\frac{1}{2}}})\, X_{1B} + (.4 + \frac{.5}{D_2^{\frac{1}{2}}})\, X_{2A} + (.4 + \frac{.5}{D_2^{\frac{1}{2}}})\, X_{2B}$$

ST: $X_{1A} + X_{2A} = 20$

$X_{1B} + X_{2B} = 5$

where:

$D_1 = X_{1A} + X_{1B}$

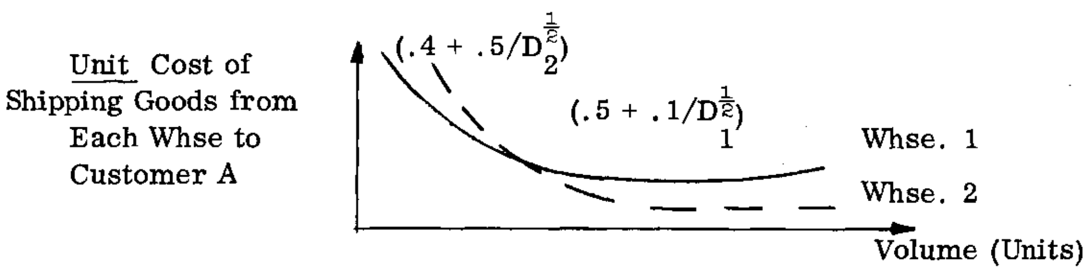
$D_2 = X_{2A} + X_{2B}$

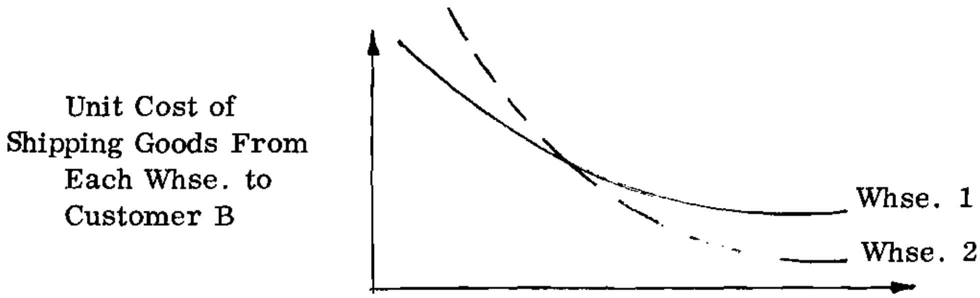
Writing the problem in terms of a standard transportation matrix,

	Customers	A	B	Slack	Capacity
Warehouses:	1	$.5 + .1/D_1^{\frac{1}{2}}$	$.5 + .1/D_1^{\frac{1}{2}}$	0	25
	2	$.4 + .5/D_2^{\frac{1}{2}}$	$.4 + .5/D_2^{\frac{1}{2}}$	0	25
Demand		20	5	25	50

R_{ij} $\qquad$ d_j

2. Graphical Interpretation of the Problem.





Notice that the cost functions "cross" (a graph of the total cost of shipments would exhibit similar characteristics), causing a least-cost warehouse in one interval not to be the least-cost warehouse in another interval.

3. Solution (Hammond's Strategy)

(1) Let $D_1 = D_2 = 4$

then,

$$R_{1,A} = .5 + \frac{.1}{(4)^{\frac{1}{2}}} = .5 + .05 = .55$$

$$R_{1,B} = .55$$

$$R_{2,A} = .4 + \frac{.5}{(4)^{\frac{1}{2}}} = .4 + .25 + .65$$

$$R_{2,B} = .65$$

(2) Based on these cost coefficients, the solution to the transportation problem will be

$$X_{1A} = d_A = 20, \; X_{2A} = 0, \; X_{1B} = 5, \; X_{2B} = 0$$

(3) Let next approximated warehouse throughputs (D_1 and D_2 in calculation

of matrix cost elements R_{1A}; R_{2A}; R_{1A}; R_{2B} be the actual throughputs resulting for the previous transportation problem solution.

Hence,

$$D_1 = X_{1A} + X_{1B} = 25$$

$$D_1 = X_{2A} + X_{2B} \cong 1 \text{ (so that we will not be dividing by "0")}$$

then, the new cost elements are

$$R_{1,A} = .5 + .1/(25)^{\frac{1}{2}} = .5 + .02 = .52$$

$$R_{1,B} = .52$$

$$R_{2,A} = .4 + .5/(1)^{\frac{1}{2}} = .90$$

$$R_{2,B} = .90$$

(4) Based on these cost coefficients the solution to the problem is

$$X_{1A} = 20,\ X_{1B} = 5,\ X_{2A} = 0,\ X_{2B} = 0$$

(5) Since the throughputs have not changed from the previous solution, the final problem solution has been reached.

$$\text{Total Cost} = .52\ (25) + .90\ (0) = 13.0$$

However, this is not the optimal problem solution since, letting

$$X_{1A} = 0,\ X_{1B} = 0,\ X_{2A} = 20,\ X_{2B} = 5$$

implies the following Total Cost

$$\text{Total Cost} = \left[.5 + \frac{.1}{(1)}\right](0) + \left[.4 + \frac{.5}{(25)^{\frac{1}{2}}}\right](25)$$

$$= 0 + \left[.50\right](25)$$

$$= 12.5$$

Therefore, Hammond's strategy <u>can</u> result in a suboptimum solution when the problem contains two or more cost functions which "cross."

LITERATURE CITED

(1) R. J. Atkins and H. J. Shriver, "New Approach to Facilities Location," Harvard Business Review, 46, 70-9 (1968).

(2) M. L. Balinski, "Integer Programming: Methods, Uses, Computation," Mgt. Sci., 12-A, No. 3, 286-93 (1965).

(3) __________, "Fixed-Cost Transportation Problems," Naval Research Logistic Quarterly, 8, 41 (1960-1961).

(4) M. A. Efroymson, and T. L. Ray, "A Branch-Bound Algorithm for Plant Location," ORSA, 14, 361 (1966).

(5) G. Sá, "Branch and Bound and Approximate Solutions to the Capacitated Plant-Location Problem," ORSA, 17, 1005 (1969).

(6) R. L. Gue, J. C. Liggett, and K. C. Cain, "Analysis of Algorithms for the Zero-One Programming Problem," Communications of the ACM, 11, 837-44 (1968).

(7) A. A. Kuehn and M. J. Hamburger, "A Heuristic Program for Locating Warehouse," Mgt. Sci., 9, 643 (1963).

(8) E. Felderman, F. A. Lehrer, and T. L. Ray, "Warehouse Location Under Continuous Economics of Scale," Mgt. Sci., 12, 670-85 (1966).

(9) H. N. Shycon and R. B. Maffei, "Simulation-Tool for Better Distribution," Harvard Business Review, (Nov.-Dec., 1960), 67-75.

(10) W. J. Baumol and P. Wolfe, "A Warehouse Location Problem," ORSA, 6, 252-63 (1958).

(11) M. L. Balinski and H. Mills, "A Warehouse Problem," prepared for: Veterans Administration; Mathematics, Princeton, New Jersey, April 1960.

(12) R. Hammond, "Reducing Fixed and Variable Cost of Distribution," Mgt. Tech., 3, 119-27 (1963).

(13) C. H. White, "Production Allocation with Set-Up and Concave Material Cost," ORSA, 17, 958 (1969).

(14) K. Spidelburg, "Algorithm for the Simple Plant-Location Problem with Side Conditions," ORSA, 17, 85-111 (1969).

(15) G. L. Nemhauser, Dynamic Programming, John Wiley and Sons, Inc., New York (1967).

(16) A. H. Bowker, and G. J. Lieberman, Engineering Statistics, Prentice-Hall, Inc., Englewood Cliffs, New Jersey (1959), p. 286.

(17) G. Hadley, Linear Programming, Addison-Wesley Publishing Co., Inc., Reading, Mass. (1962).

OTHER REFERENCES

Adams, T. H., "Use of Linear Programming in the Selection of Optimum Shipping Points," Transportation J., 8, 1-20 (Spring, 1969).

Bellman, R., "An Application of Dynamic Programming to Location-Allocation Problems," SIAM Review, 7, No. 1, 126 (1965).

Bergendahl, G., "Separable Programming Methods for Determining Plant Location Under Economics of Scale," Analysis of Business Systems, G. Fisk (ED.), CWK Gleerup Publishers, Lund, Sweeden, 95-114 (1967).

Cooper, L., "Heuristic Methods for Location-Allocation Problems," SIAM 6, 37 (1964).

__________, "Location-Allocation Problems, ORSA, 11, 331 (May-June, 1963).

Curry, G. L., et. al., "A Dynamic Programming Algorithm for Facilities Location and Allocation," AIEE Transactions, 1, no. 2, 133 (1969).

Dwyer, P. S., "Solution of the Fixed Charge Transportation by Reduced Matrixes," Naval Research Logistics Quarterly, 13, 289-313 (September, 1966).

Enehorg, C. G., "How to Plot the Ideal Location for a Warehouse," Mgt. Methods, 13, 52 (January, 1958).

Francis, R. L., "On the Location of Multiple New Facilities with Respect to Existing Facilities," Journal of Industrial Engineering, 15, no. 2, 106 (1964).

Fredericks, W. A., and Smykay, E. W., "An Index Based Method for Evaluating Warehouse Locations," Transportation Journal, 3-4, 30 (1963).

Gross, D., and Soriano, A., "The Effects of Reducing Leadtime on Inventory Levels--Simulation Analysis," Mgt. Sci., 16, no. 2, 3-61 (1969).

Manne, A. S., "Plant Location Under Economics of Scale--Decentralization and Computation," Mgt. Sci., 11 (November, 1964).

__________, Investment For Capacity Expansion, Size, Location, MIT Press, Cambridge, Mass. (1967).

Meyer, C. F., "Model for Production-Distribution Networks with Transit Billings," Mgt. Sci., 14, B204 (1967).

Parsons, J. A., "Branch and Bound Algorithms: The Plant Location Problem," Journal of Systems Mgt., 20, 43-45 (October, 1969).

Polopolus, L., "Optimum Plant Numbers and Location for Multiple Product Processing," Journal of Farm Economics, 47, 287 (May, 1965).

Rivera, L., "The Systems Approach to Warehouse Planning," Thesis, Georgia Institute of Technology, (1968).

Stollsteimer, J., "A Working Model for Plant Numbers and Locations," Journal of Farm Economics, 45, 631 (August, 1963).

Sussans, J. E., "Some Problems Associated with the Distribution of Consumer Products," O. R. Quarterly, 19, 161 (June, 1968).

Zangwill, W. I., "A Deterministic Multi-Product, Multi-Facility Production and Inventory Model," ORSA, 14, 486-507 (1966).

____________, "The Piecewise Concave Function," Mgt. Sci., 13, 900 (July, 1967).